№ 5573.1.9

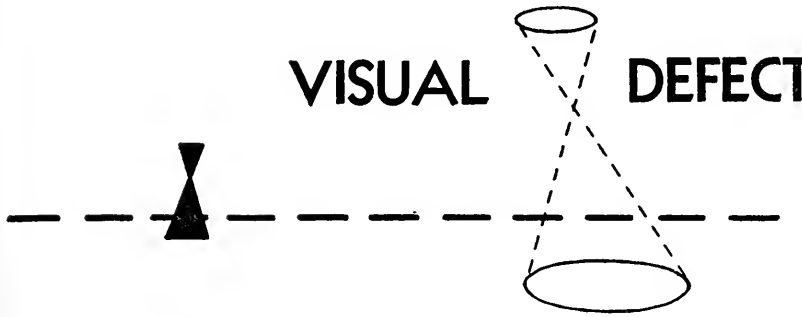
335.97

GIVEN BY

45
*5573.147
no. 345

SCREENING SCHOOL CHILDREN for VISUAL DEFECTS

SCREENING SCHOOL CHILDREN for VISUAL DEFECTS



report of a study conducted
in St. Louis, Missouri, 1948–49

Marian M. Crane, M.D., Franklin M. Foote, M.D.,
Richard G. Scobee, M.D., and Earl L. Green, Ph.D.,
with statistical assistance of Bronson Price, Ph.D.

This study was carried out with the joint support of the National Society for the Prevention of Blindness, the Children's Bureau of the United States Department of Health, Education, and Welfare, the Division of Health of the Missouri State Department of Public Health and Welfare, the St. Louis Board of Education, the Department of Ophthalmology, Washington University School of Medicine, and the Office of Naval Research.

Dr. Crane is Chief of the Research Interpretation Branch of the Children's Bureau. Dr. Foote is Executive Director of the National Society for the Prevention of Blindness. Dr. Scobee (deceased) was Assistant Professor of Ophthalmology, Washington University Medical School. Dr. Green, at the time of the study, was Associate Professor of Zoology, Ohio State University, and he is now Geneticist, Biology Branch, Division of Biology and Medicine, of the United States Atomic Energy Commission. Dr. Price is Public Welfare Research Analyst, Division of Research, Children's Bureau.

CONTENTS

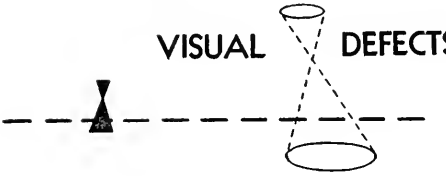
	Page
SUMMARY AND CONCLUSIONS	1
SUBJECTS AND METHODS	5
Subjects	5
Screening Procedures Studied	6
The Testers and Their Training	7
Scheduling of Tests and Examination	8
Testing Conditions and Supervision of Testers	8
Procedure for the Clinical Examination	9
Methods Used in Screening Procedures	11
Standards for Referral	12
FINDINGS OF THE STUDY	13
Clinical Findings	14
Distributions of Measurements	16
Screening Efficiency	16
Relative Efficiency of the Procedures	22
Relative Efficiency of Testers	24
Variations in Total Referrals	25
Effect of Retesting Before Referral	27
Reliability and Validity of Component Parts of Screening Procedures	28
Reliability Coefficients	28
Validity Coefficients	33
Contributions of Specific Parts of Tests to Total Tests	36
Factors That Might Influence Test Results	38
Tests of Acuity without and with Occlusion of Opposite Eye	38
Errors in Recording Test Results	40
Learning Factor	42
Effect of Length of School Experience	44
Influence of Economic Status	45
DISCUSSION	47
Why Is Screening Efficiency Not Higher?	49
Improving Test Construction	50
More Effective Use of Present Tests	53
APPENDIX A. History and Personnel of the Study	57
APPENDIX B. Standards for Referral	63
APPENDIX C. Distributions of Measurements	70
APPENDIX D. Discussion of Correlations	89

SCREENING

SCHOOL CHILDREN

for

VISUAL DEFECTS



SUMMARY AND CONCLUSIONS

A STUDY to determine the efficiency of certain procedures commonly used or recommended for use in schools to screen children for visual defects was conducted in St. Louis in 1948 and 1949.

Six hundred and nine sixth-grade students and 606 first-grade students in the public schools were given a complete ophthalmological examination and tested with certain vision screening procedures.

The screening procedures studied were: Teacher Judgment, the Snellen Test, a Near Vision Test using the Lebensohn or Guibor charts, the Massachusetts Vision Test, the Keystone View Company Telebinocular Test, combinations of some of these procedures, and, for sixth-grade students, 2 procedures developed for use in industry rather than in schools: the Bausch and Lomb Ortho-Rater and the Americal Optical Company Sight-Screener.

Each screening procedure, except Teacher Judgment, was administered twice to every student—once by the school nurse for the school attended by the student, and once by a technician who tested all students in the study. The twelve school nurses received training in testing technique believed to correspond to what is usually considered adequate preparation; the training given the technician was more extensive than that given the nurses and he acquired more experience in testing. The Snellen Test was given each student by 3 testers, the technician, the school nurse, and the student's classroom teacher. The teachers received only brief training in administration of the Snellen test.

The ophthalmologist found 31 percent of sixth-grade students and 23 percent of those in first grade to need referral for professional eye care, or 27 percent of the entire group of 1,215 students.

The ophthalmologist's judgment as to need for referral was used as the criterion against which to evaluate efficiency of the vision screening procedures.

The referrals and nonreferrals obtained with each screening procedure are presented (table 2a) in a manner that shows also whether or not they were referrals by the ophthalmologist.

The best measure of the efficiency of a screening procedure is the correlation of its results with the ophthalmologist's findings; the actual proportions of correct and incorrect referrals are determined by two factors, the correlation and the standard for referral.

None of the correlations obtained show high screening efficiency. For the sixth grade the Snellen and Massachusetts Vision Tests gave average correlations of approximately 0.45. Next in order are the Ortho-Rater and Sight-Screener, with average correlations of 0.33, but it is doubtful whether practical or statistical significance attaches to the margin they have over the Telebinocular and Near Vision tests, for which the correlations are of the order of 0.27 and 0.29. For first grade the pattern of findings is similar.

In general, with the cut-off points used, the larger the number of component tests in a procedure, the higher the proportion of total referrals. This means that the multiple-test procedures gave more correct referrals, but in the absence of greater screening efficiency they also gave a higher proportion of over-referrals.

Teacher Judgment, as obtained in the study, correlated only about 0.16 with the ophthalmologist's judgment; combining Teacher Judgment with other screening tests decreased the efficiency of these other tests. Procedures that included Teacher Judgment did not show the usual relationship between total referrals and number of component tests, since Teacher Judgment alone had an unusually high referral rate. It is possible that a better method of obtaining Teacher Judgment, with more extended observation and with referral only after teacher-nurse conference, would make a greater contribution to screening efficiency.

No marked improvement in screening efficiency is found for the combination of a Near Vision Test with the Snellen over that for the Snellen alone, although the total number of referrals can be increased by using the combination of tests.

Repeating a screening procedure and referring only the students who fail both times gives some gain in screening efficiency, but a single repetition of the procedure as a whole cannot be relied upon as an effective remedy for low screening efficiency.

The correlation with clinical judgment of the screening results obtained by the technician, the tester with more extensive training and experience, was, on the average, slightly better than that of the results obtained by the nurses, but the difference was not great. The teachers were apparently able to administer the Snellen Test about as efficiently as the technician and nurses.

Examination of the effects of certain factors that might influence screening test results led to the following conclusions:

Students who obtained low scores on Ortho-Rater or Telebinocular subtests of monocular acuity tended to obtain somewhat higher scores when retested with occlusion of the opposite eye, but it is not clear to what extent the tendency resulted from a learning factor, and it was not observed in comparable testing with the Sight-Screener.

The technician made fewer errors in recording screening-test scores than the nurses, especially when the latter were just beginning their testing programs, demonstrating another aspect of the value of experience for testing efficiency.

There is evidence that at the first-grade level a learning factor may have resulted in slightly higher scores the second or third time a test was administered to a student than the first time, especially if the test was a difficult one for the student and was administered the first time by the less experienced tester, but it is unlikely that learning factors materially affected overall results regarding screening efficiency.

Comparison of test scores obtained in the first and second semesters of the first grade show such slight differences that the findings from both semesters may be considered applicable to students entering the first grade.

There is no evidence of any consistent variation according to economic status in the proportions of students referred by the ophthalmologist or the proportions who obtained low scores on subtests, but a higher proportion of Negroes was referred by the ophthalmologist than of white students, irrespective of economic status.

Referral level and relative screening efficiency of a procedure may shift with any changes in the methods of administering or scoring the tests. The contribution made by this study lies not merely in ranking the procedures as now administered according to their usefulness for screening purposes, but also in what it tells as to why screening efficiency is not higher and as to how it can be improved.

The efficiency of a screening procedure depends upon the reliability and validity of its component parts. Examination of test-retest correlations and validity data shows that several of the screening procedures include component tests having such low reliability of measurement, at least as used with school children, that a high degree of correlation with clinical judgment is not to be expected from the screening procedure as a whole. With first-grade students, only the tests of visual acuity at far point show high enough correlation with corresponding clinical tests to indicate efficiency for screening purposes.

There are recognized methods of increasing the reliability of testing procedures, but these require devotion of more time to administration of the tests than is customarily allowed in screening programs. Hope that, even with the most ingenious instrument devisable, any visual function

can be measured reliably by a quick check holds practically no promise of success, especially with children.

From the administrative point of view the amount of time required for a screening procedure is an important consideration. But in screening with a procedure that consists of a battery of tests, time for obtaining more reliable scores on some of the component tests can be gained by eliminating from the battery other tests that contribute little to the overall efficiency of the procedure. Moreover, there may well be compensation for additional time given to screening tests in the saving of time that nurse or teacher would otherwise devote to follow-up of incorrect referrals.

Further attention to certain recognized principles of test construction should enable those concerned with improvement of vision screening procedures to develop methods of administering the tests that will result in greater reliability of test scores and, hence, in better screening efficiency.

On the basis of the study findings certain suggestions are offered to school health administrators for use of testing materials as now set up. Of primary importance are good testing conditions, an unhurried testing schedule, and a definite plan for repetition of tests so as to obtain reliable scores for all component subtests. If facilities for follow-up of referrals are too limited to reach more students than will be referred by the Snellen Test (or other test of visual acuity at far point), there is probably nothing to be gained by using a screening procedure with a higher referral rate. Where there are more adequate follow-up facilities, one of the multiple-test procedures may be preferred in order to find more of the students who need care; as now set up the Massachusetts Vision Test is the most efficient of these, but all could be made more efficient by following the suggested modifications. It is unlikely that anything is to be gained by using procedures other than the Snellen, or possibly the Massachusetts Vision Test, below the third or fourth grade.

SUBJECTS AND METHODS

THE PURPOSE of a program for screening school children for visual defects is to find those students who need treatment or observation by an eye specialist. The final arbiter of whether or not treatment is needed is the eye specialist.

The purpose of this study is to determine the efficiency of certain procedures commonly used or recommended for use in school screening programs as measured by their success in identifying those students who are shown by ophthalmological examination to be in need of treatment.

For practical usefulness a screening procedure should be one that can be administered successfully without elaborate preliminary training of the tester and within a relatively brief period of time for the individual test. Secondary purposes of this study are, therefore, to obtain information as to the amount of preliminary training in testing technique needed for successful administration of the tests and as to the approximate amount of time required for each testing procedure.[1]

Subjects

Subjects of the study were 609 sixth-grade students and 606 first-grade students in St. Louis public schools. The schools in which the study was conducted were selected to give a rough cross-section of socio-economic groups. Parents of all students in the appropriate grades in these schools were informed about the study and asked to give permission for the ophthalmological examination. All students whose parents gave this permission were included in the study except those who were ill or absent for other reasons at the time of the testing program in that school. No record was kept of the reasons for, or number of parental refusals. It was the impression of the investigators that the refusals amounted to less than 10 percent. and that very few were for reasons connected with the eyes. It is possible, however, that the frequencies of the visual defects found by the ophthalmologist might have been slightly different if all students in the two grades could have been examined.

Of the sixth-grade students, 474 were white and 135 were Negro; 303

[1] Results of the investigation of testing time have been reported in "Study of Procedures used for Screening Elementary School Children for Visual Defects," by Marian M. Crane, Richard G. Scobee, Franklin M. Foote, and Earl L. Green. American J. Pub. Health 42:1430–1439, Nov. 1952; and The Sight Saving Review 22:141–153, 1952.

287418P—54——2

were boys and 306 girls. In the first grade there were 539 white students and 67 Negroes, 286 boys and 320 girls. The distribution according to age at last birthday was as follows:

Sixth grade		First grade	
Age	Number of students	Age	Number of students
Years		*Years*	
10	38	5	21
11	267	6	447
12	166	7	121
13	81	8	13
14	37	9	4
15–18	20		

Sixth-grade students were selected for study because they were old enough to give full cooperation. First-grade students were studied to determine whether beginning readers would give the cooperation necessary for satisfactory testing.

Screening Procedures Studied

Seven procedures were selected for evaluation. These were:

Teacher Judgment: A judgment by the classroom teacher as to whether or not the student should be referred for an eye examination, based on her observation of the student.

Snellen Test; Massachusetts Vision Test; Keystone View Company Telebinocular Test: These 3 procedures are widely used in school health programs and are representative of all the tests commonly used for visual screening of school children.

American Optical Company Sight-Screener; Bausch and Lomb Ortho-Rater. These 2 procedures, developed primarily for testing the vision of industrial workers, were studied to determine their suitability for use with children of school age. Since they are designed for subjects who have more ability to read than is to be expected in the first grade, they were studied with sixth-grade students only.

Near Vision Test. A separate test of visual acuity at reading distance was included because some school health authorities believe that a combination of such a test with a Snellen Test would be a more efficient screening procedure than the Snellen alone, while others argue that a brief test of near vision has little value because the child has enough power of accom-

modation to be able to read the test lines even though he could not read comfortably for a longer period.

Combinations of Procedures. In addition to the study of these 7 procedures, the effect of combining certain procedures was examined. The combinations studied are: Teacher Judgment and Snellen Test; Snellen Test and Near Vision Test; Teacher Judgment, Snellen, and Near Vision; Teacher Judgment and Massachusetts Vision Test.

The Testers and Their Training

To obtain information as to the amount of preliminary training in testing technique required for successful administration of the screening procedures, 3 different persons tested each student.

One *technician* tested every student in the study with all of the procedures, excepting, of course, Teacher Judgment. This technician received thorough training. For the Massachusetts Vision Test, Ortho-Rater, Sight-Screener, and Telebinocular Tests he was trained by persons who had participated in developing the tests or other representative of the manufacturer. In the course of testing more than 1,200 students in the study, he acquired considerable experience with the tests.

The *nurse* for each school in which the program was being conducted gave to all of the students studied in that school the same tests that the technician gave. Twelve different nurses served the 14 schools, so the term "nurse" as used in this report actually represents a group of 12 nurses. The nurses were taught testing procedure by the "nurse-coordinator," who had received the same preliminary training as the technician.

A prescribed plan was followed in training the nurses. One full school day was given to instruction and practice in each procedure except the Near Vision Test, for which less time was needed. Half a day was devoted to review and to learning other details of the program. The consultants who trained the technician and nurse-coordinator agreed that this plan gave the nurses a preparation they would consider adequate and comparable to that usually given to the tester in a school program, although it was less extensive than the training given the technician. The procedures were taught to each nurse in a different sequence. The sequence to be followed for any one nurse was determined by lot.

Each student's classroom teacher gave him a Snellen Test. In some school health programs it is customary to have each classroom teacher administer this test to her own students, usually after a very limited amount of training in testing procedure. In order to determine how effectively teachers with such limited training can administer the Snellen Test, the teachers who participated in the testing were given only a brief preparation.

The training of these 3 types of testers was planned to represent the training it would be practical to provide in each of 3 different situations:

(a) in a program that was preparing one tester to test a number of schools; (b) in a program using one tester for each school; or (c) in training every classroom teacher.

Scheduling of Tests and Examinations

To minimize the possible influence of the learning factor on the student's performance in tests or clinical examination, the sequence in which the examination and the various tests were given was randomized. Each student was assigned by chance to one of many possible schedules. For half of each age group the clinical examination preceded the screening tests, for the other half all screening tests were completed before the clinical examination. Each of the 6 possible sequences in which teacher, nurse, and technician might test a student was followed for one-sixth of the students. The order in which the technician and nurse administered the different screening tests was determined by chance, separately for each tester.

There were altogether for sixth-grade students 1,728 possible sequences of clinical examination and screening tests. For the first-grade students there were fewer tests and consequently fewer testing sequences.

Students scheduled for the screening tests first were given the ophthalmological examination as soon as possible after being tested, usually within 1 to 4 days. For those who had the clinical examination first, at least 7 days were allowed to intervene before other testing in order that all effects of the cycloplegia might disappear.

Students were scheduled for testing by the nurse and technician in groups of 2 or 3—usually 2 for sixth-grade students and 3 for first-grade students. The tester took the students in the group alternately, so that after each test a student had a rest while the other student or students in his group received a test. During the waiting interval the students were kept in the same room—with the same amount of lighting—in which they were tested and were kept entertained with occupations that did not require close use of the eyes.

About 80 minutes were required for the nurse or technician to give all the tests to 2 sixth-grade students or 3 first-grade students. Daily testing schedules were therefore set up in 3 blocks of time for testing by technician and nurse, with each tester scheduled for 2 groups of students in the morning and 1 in the afternoon. Testing periods for the teacher were interspersed as necessary to make her the first, second, or third tester.

A student tested by the nurse during the first morning period was scheduled for the technician during the afternoon, and vice versa, so there was a rest period of more than 2 hours between the 2 testing periods. Those tested for the first time during the second morning period went to the other tester during the second period on the following day. This plan had to be modified at times to meet special situations but it was only rarely that it was necessary to schedule a student for his second series of tests immediately after the first series.

Testing Conditions and Supervision of Testers

All screening tests were given in a single room or in 2 adjoining rooms. Testing equipment was set up in advance with special attention to

lighting and to space—avoiding glare and reflections, maintaining the correct 20-foot alley for the Massachusetts Vision Test, and avoiding too close proximity of the different tests.

In a school vision-screening program the tester usually works without supervision. If he fails to follow correct testing procedure there is no one at hand to correct him, although he can usually get advice if he is uncertain about the procedure. Since the purpose in using testers with limited training in this study was to find out how successfully a tester with such training could administer the tests, their testing was not supervised. If a member of the staff noticed an error in the nurse's or teacher's testing procedure, no comment was made. But if the tester asked questions or sought advice she was given any help she wanted. Actually very few errors in testing procedure were observed.

Testers were checked, however, on the completeness of their records, since it was important that no record should have to be discarded because of incomplete recording. Whenever an entry was omitted or ambiguous it was called to the tester's attention at the end of the day. Often the intended entry was obvious and the correction could be made without further testing. But if there was any question as to what the correct entry should be, the student was recalled and the part of the test involved was repeated. Since errors of this kind reflect the efficiency or inefficiency of the tester and, if they are not caught, may result in incorrect classification of the student in relation to need for referral, a record was kept of the number of errors each tester was required to correct.

Procedure for the Clinical Examination

The clinical examination was conducted during the middle of the day and was begun shortly after the children had had their lunches. It was given in the out-patient department of the McMillan Hospital, to which the students were transported by taxicab, with an adult escort. A waiting room and the rooms in which the examinations were given were used for no other purpose while the students were there.

Three examiners were used and each performed the same one-third of the clinical examination on each child.

A child was sent first to examiner A. A brief ophthalmologic history was recorded with specific inquiries made about the child's ability to see the blackboard clearly, the presence of blurred vision at any time, drowsiness on reading, and headaches—particularly when watching moving objects such as motion pictures. The presence or absence of these specific symptoms was noted on the blank. (A history obtained only from the child has obvious limitations, but was the best available.) An external examination was then performed and this consisted of an inspection of the lids, lashes, conjunctiva, cornea, iris, and lens. The next step was the determination of the child's visual acuity at 20 feet, with glasses if

worn, and then without them, each eye being tested separately. The right eye was usually tested first. If the eye first tested appeared to have subnormal acuity, at the completion of the test it was retested and frequently found to be normal. If glasses were worn, their prescription was determined from a trial case nearby and this was also recorded. The child next went to examiner B.

Examiner B's activities were limited to a study of muscle balance. A cover test with loose prisms was first performed at far and then at near and the findings recorded. A Maddox rod test for heterophoria was next done, except in those children who had heterotropia and in whom it could not be performed. The Maddox rod was mounted on a Stevens phorometer and was always placed before the right eye. Measurements were made for far (20 feet) and then for near (13 inches). A Maddox wing test was performed at 13 inches. Prism vergences were studied next in the following order: (1) prism convergence at far, (2) prism convergence at near, (3) prism divergence at far, and (4) prism divergence at near. A Stevens phorometer with a Risley rotary prism was employed for vergence determinations. The near point of convergence was then measured and the second portion of the examination was completed.

The children were taken in a group to a separate room where a nurse instilled eye drops. These were 2 percent homatropine in 1/5000 zephiran. A drop was instilled in each eye of each child every 10 minutes for 5 instillations. Ten minutes after the last instillation of homatropine, the children were taken to examiner C. Retinoscopy was performed on each eye and the findings recorded. The indicated lenses were placed in the trial frame and visual acuity determined. If a visual acuity of 20/20 in each eye was found at this time, no further subjective refinement of the refraction was attempted. If 20/20 was not attained, efforts were continued until it could be, or until it became obvious that it could not be under these circumstances. The fundi were then examined carefully and any findings noted. Examiner C was prepared to perform visual field studies—both central and peripheral—if these were indicated. At the conclusion of the study, it was found that visual field determinations had been made upon only one child (because of a choked disc).

One-half percent pilocarpine was instilled once into each eye of each child at the conclusion of examiner C's study. This completed the clinical examination.

It is to be emphasized again that each of the 3 examiners performed the same portion of the examination on every child. Ideally, of course, a single examiner would have performed the entire examination. Three examiners were necessary because the time available to each examiner was limited. It was nevertheless hoped that any examiner-variable would be kept to a minimum with the procedure outlined.

Examiner B reviewed all of the findings of the examination and recorded a clinical judgment: "refer" if the student needed treatment or observation, "nonrefer" if treatment was not needed.

Methods Used in Screening Procedures

Teacher Judgment was based on the classroom teacher's observation of the student. To alert the teachers to the possible significance of signs or complaints that are commonly associated with eye trouble in children, a physician or nurse on the study staff met with the teachers in each school shortly before the testing program began in that school and reviewed these points with them. The teacher recorded her judgments before she did any testing: "refer" if she thought there was evidence that the student needed to have his eyes examined, "nonrefer" if she did not think an eye examination was indicated.

The **Snellen Test** was given with the symbol E lines on the Massachusetts Vision Test chart. The E symbols on this chart are constructed according to the principle of the Snellen Chart, so it was assumed that a test with these lines would be equivalent to a test with the corresponding lines on a properly placed and correctly lighted Snellen chart. Since the first part of the Massachusetts Vision Test measures the subject's ability to read the 20/20 and 20/30 lines, the results obtained by the technician and nurse on this part of the Massachusetts Test are used as the Snellen Test measurements by these 2 testers. Only the teacher gave the Snellen as a separate test. She was instructed to follow the testing procedure used in the Massachusetts Test with one addition: if a student failed to read either the 20/20 or 20/30 line, the 20/40 line on the chart was to be exposed and the student's ability to read this line was to be tested.

Before any testing of first-grade students, their teacher gave them a brief practice drill in the classroom in the use of their hands to show the position of the E symbol as it appeared on a card held up before them.

The **Massachusetts Vision Test** was administered according to the procedure recommended by the Massachusetts State Department of Public Health, except that, instead of discontinuing the test when a student failed any part of it, the complete test was given to every student.

According to this procedure the first part of the test, reading the E symbols, is invariably started with the 20/20 line. A child who hesitates to give an answer is urged to guess. If the tester feels that a failing performance can be bettered, as many as 3 consecutive attempts at the same level may be made. The best single reading performance is then taken as the final score. Repetition is always for an entire line, never for one symbol only. In repeating a line the student is asked to read the corresponding line in the other box, or the order in which the symbols are to be read is varied, as by requesting the child to read from right to left, Guessing, urging, and confirmation are used especially in the testing of young children.

In both the first and second parts of the Massachusetts Vision Test a student is considered to have recognized a line if he described correctly the position of 4 or more of the 6 symbols.

The **Ortho-Rater, Sight-Screener,** and **Telebinocular** tests were administered according to the instructions provided by their manufacturers.

It should be noted that the Ortho-Rater and the Sight Screener were developed for use with adults rather than children. The tests of color vision included in these tests were omitted from the study because color-blindness is not a cause for referral for eye care, though color-vision was tested with one of the instruments in order to obtain data requested by the manufacturer.

For the **Near Vision Test** students were asked to read 3 lines from a chart placed 14 inches from the eyes. For the sixth grade, the lines consisted of 6 letters each from the 14/14, 14/17, and 14/23 lines of the Lebensohn Chart. Recognition of 4 letters was accepted as satisfactory performance. First-grade students were asked to indicate the position—right, left, up or down—of 4 symbols each in the 14/14, 14/17, and 14/21 lines of the Guibor Chart. A line was considered to have been read correctly if the student recognized the position of 3 of the 4 symbols.

In studying **combinations of procedures,** a student is considered to have been referred by the combination if he is referred by any one of the constituent procedures.

Standards for Referral

For the purposes of this study it was necessary to define "limits of normal" for the measurements made in the clinical examination, and standards for referral for the screening procedures. The methods by which these definitions were made, and the standards that were adopted, are presented in appendix B.

FINDINGS OF THE STUDY

THE FULL COOPERATION given by all those who participated in the study made it possible to carry out the testing and clinical examinations almost exactly as planned. Only two problems that arose call for special comment.

The ophthalmologists found considerable difficulty in administering the clinical examination to Negro first-grade students. It was frequently impossible to elicit from a young Negro child any response to questions as to what he was seeing. The difference in the cooperation given by Negro first-grade students from that given by white children of the same age group was attributed to their being more ill at ease. The experience of a taxi ride to the hospital building, escorted by an unfamiliar adult, and the strange surroundings of the waiting room and examination rooms were probably novel enough to be disturbing to many first-graders, both white and Negro, but to a larger proportion of the Negro children. In addition, the Negro children had to adjust to being escorted and examined by white people. In their home setting and in their school, where they have Negro teachers, nurse and physician, many of these children had probably had little previous contact with white people. A certain lack of poise under these conditions is entirely understandable.

If it had been possible to have the clinical examination of the Negro children conducted by Negro physicians and nurses this difficulty might have disappeared. Since this was not possible, the number of first-grade Negroes included in the study was kept relatively small.

One other difficulty was related to the method of obtaining Teacher Judgment. As the study proceeded those directing it became increasingly aware that to ask a classroom teacher to record at any one time judgments for all of her students as to whether or not they needed an eye examination was not the same thing as obtaining the benefit of the teacher's observation of her students over a longer period. With general interest aroused in the study, the teachers were eager to cooperate and probably overly ready to interpret any deviation from usual behavior as a sign of eye trouble. Under more usual conditions the teacher would be alert to complaints or behavior suggestive of eye trouble but would not be making a special search for them. She would report any such observations to the school nurse or physician who would make the final decision as to referral. The differences between Teacher Judgment as obtained in the study and teacher observation under more usual circumstances were unavoidable but should not be forgotten in evaluating the results obtained.

Clinical Findings

Because of the death of the study's Ophthalmological Director, it is not possible to include in this report as complete an analysis and interpretation of the clinical findings as he had hoped to present.

Of the 1,215 students examined, 327 were found by the ophthalmologist to need treatment or observation because of their eyes. These included 190, or 31 percent, of the sixth-grade students, and 137, or 23 percent, of the first-grade students. In the sixth grade 29 percent (137) of the white students and 39 percent (53) of the Negroes needed treatment. In the first grade, treatment was indicated for 22 percent (119) of white students and 27 percent (18) of Negro students.

Of the 609 sixth-grade students, 42, or 7 percent, were wearing glasses. Approximately 8 percent of the white students had glasses, and about 4 percent of the Negro students.

The glasses worn by sixth-grade students averaged a one-half diopter correction for hypermetropia, with no correction for astigmatism. At this age the actual need for a half-diopter correction for hypermetropia is very small. This suggests that many of the children who had received an eye examination had been given glasses when no glasses were needed. As a result of his examinations in the study the ophthalmologist concluded that 31 of the students with glasses either did not need glasses or needed different glasses.

Of the 606 first-grade students, only 7 were wearing glasses. All 7 were white children.

The eye conditions the ophthalmologist found, for which treatment was needed, are shown in table 1.

Of the entire group, 149 students, or 12.3 percent, had hypermetropia, hypermetropic astigmatism, or both. One hundred, 8.2 percent, had myopia, myopic astigmatism, or both. Fifty-three, or 4.4 percent, needed treatment because of muscle imbalance only. Of these, 24 had exotropia or esotropia, 24 had lateral heterophoria of significant degree, 4 needed treatment for hyperphoria, and 1 had poor prism divergence.

Of the 8 students who had inflammatory conditions of the eyes demanding attention, 4 had conjunctivitis or pink eye, 2 had a Meibomian abscess on one lid, 1 had severe squamous blepharitis and 1 had episcleritis. Six children had conditions that are classed as miscellaneous. One of these had a sunken, sightless eyeball that should be removed and replaced with prosthesis; 1 child had nits (lice) in her eyelashes; one had a congenital cataract in one eye that should be removed; 1 had papilledema (choked disc) in one eye; 2 had subnormal visual acuity that was not improved by lenses.

The ophthalmologist applied a high standard in selecting the students needing treatment in order that no child needing care even slightly would be missed. The group includes the children who for the time being needed

TABLE 1

Conditions Requiring Treatment Found by Ophthalmological Examination of 609 Sixth-Grade and 606 First-Grade Students

Condition		Students	
		Number	Percent
Total students needing treatment		327	26.9
Hypermetropia, hypermetropic astigmatism, or both		149	12.3
Myopia, myopic astigmatism, or both		100	8.2
Muscle imbalance only		53	4.4
Esotropia	18		
Exotropia	6		
Esophoria	7		
Exophoria	17		
Hyperphoria	4		
Poor prism divergence	1		
Ocular inflammation		8	.7
Miscellaneous		6	.5
Not recorded[1]		11	.9

[1] At an early stage in the analysis of the study data the ophthalmologist classified the clinical referrals according to the conditions requiring treatment. Later he reviewed for consistency the clinical judgments recorded on all borderline cases and transferred 11 students from the nonreferral group to the referral group. All of these students had either refractive errors or muscle imbalance, but the records are not clear as to how they should be assigned to the 3 categories of such conditions and they have not been distributed arbitrarily. They are included as referrals in the cross-tabulations against screening results; their exclusion would have had no important effect on the degree of correspondence shown.

only to be under the observation of an eye specialist as well as those who needed glasses or other active treatment. Since there is no general agreement among ophthalmologists as to the degree of refractive error or muscle imbalance that calls for correction, not all would agree on the need for treatment for some of these children. But the urgency of treatment is not necessarily proportional to the magnitude of the refractive error. A good school health program should aim at placing under professional care all students who have any real need for such care.

Among the students found by the ophthalmologist to need care were, however, some with conditions that screening tests cannot be expected to identify. Five sixth-grade students had incipient myopia but normal visual acuity. These 5 complained of headaches, so they might be identified by a screening program that included teacher observation, but not by a screening test.

Likewise teacher observation should refer the 6 children with acute

conjunctivitis or Meibomian abcess, but they might be missed by a test. This brings to 11 the number that a screening test alone could not be expected to identify, 3 percent of all those who need treatment.

Distributions of Measurements

The distributions of measurements obtained on each part of the clinical examination and the technician's screening tests are presented in appendix C. The distributions shown for sixth grade are based on measurements of both white and Negro students, those for first grade are based on measurements of white students only.

Screening Efficiency

The main findings of the study are presented in tables 2a and 2b. The data on each screening procedure vis-a-vis ophthalmologist's judgment are shown in table 2a, while table 2b gives certain statistics summarizing the results.

In table 2a the students whom the ophthalmologist classed as referrals—in need of eye care—are shown as the first figure (under R) for each grade. This figure is 190 for the sixth grade, amounting to 31 percent of the students. The corresponding figure is 137, or 23 percent, for the first grade. Alongside these figures (under N) are the remaining students in each grade, or those whom the ophthalmologist classed as nonreferrals.

The next row of the table shows that Teacher Judgment referred 80 of the sixth-grade students who actually needed care, and 117 who did not. The first of these figures may be termed "correct referrals," and the second figure "incorrect referrals" or "over-referrals". To save space the table does not show the sum of these figures (197), but obviously it is the total number of referrals made by this screening procedure.

Likewise the sum of the first two figures in the next line (110 and 302) is the total number of nonreferrals. The number 110 represents the students "missed" by Teacher Judgment, or those additional students who would have been referred with an ideal screening procedure, and 302 is the remaining number of students, or those whom neither ophthalmologist nor teacher classed as referrals. The four numbers arranged as shown in the table comprise what is called the "two-way distribution" or "2 x 2 scatter" for the screening procedure vs. the criterion.

It is difficult to judge the relative efficiency of the screening procedures from these scatters. The problem can be illustrated by considering the results obtained by the technician in testing sixth-grade students with the Ortho-Rater, the Snellen Test, high standard, and the Near Vision Test, high standard. For convenience in discussion they will be referred to as

TABLE 2 a

Referrals and Nonreferrals by Screening Procedures Versus Ophthalmologist's Judgment

"R" means students referred and "N" means not referred by a given procedure. Thus, of 609 sixth-grade students, 190 were classed as referrals by the criterion procedure (ophthalmologist's examination) and 419 were classed as nonreferrals. Of the 190, Teacher's Judgment referred 80 students ("correct referrals") and did not refer 110 students ("missed cases"). Of the 419 who were nonreferrals by the criterion, Teacher's Judgment referred 117 students ("over-referrals"), and the remaining 302 cases were not referred by either procedure.

Procedure and Tester		NUMBERS				PERCENT [1]			
		Sixth grade (total= 609)		First grade (total= 606)		Sixth grade (total= 100)		First grade (total= 100)	
CRITERION—OPHTHALMOLOGIST'S JUDGMENT from CLINICAL EXAMINATION		R 190	N 419	R 137	N 469	R 31	N 69	R 23	N 77
TEACHER'S JUDGMENT—	R	80	117	58	118	13	19	10	19
	N	110	302	79	351	18	50	13	58
SNELLEN, high std., by—									
Technician	R	85	14	64	41	14	2	11	7
	N	105	405	73	428	17	67	12	70
Nurse	R	87	34	70	51	14	6	12	8
	N	103	385	67	418	17	63	11	69
Teacher	R	91	39	81	53	15	6	13	9
	N	99	380	56	416	16	63	10	68
SNELLEN, low std., by—									
Technician	R	43	4	24	4	7	1	4	1
	N	147	415	113	465	24	68	19	76
Nurse	R	39	5	33	1	6	1	5	0
	N	151	414	104	468	25	68	18	77
Teacher	R	48	9	27	6	8	2	4	1
	N	142	410	110	463	23	67	19	76
MASSACHUSETTS, by—									
Technician	R	120	65	92	96	20	11	15	16
	N	70	354	45	373	11	58	8	61
Nurse	R	112	72	89	99	18	12	15	16
	N	78	347	48	370	13	57	8	61

[1] These figures are shown in whole percentage units in order to facilitate inspection of the scatters, and some of the figures are necessarily "forced" more than half a percentage point to make each scatter total 100. The percentages have not been used for the correlations in table 2b, which were computed from the numbers of students rather than the percentages.

TABLE 2 a—Continued

Procedure and Tester	NUMBERS				PERCENT			
	Sixth grade (total= 609)		First grade (total= 606)		Sixth grade (total= 100)		First grade (total= 100)	
CRITERION—OPHTHALMOLOGIST'S JUDGMENT from CLINICAL EXAMINATION..	R 190	N 419	R 137	N 469	R 31	N 69	R 23	N 77
ORTHO-RATER, by—								
Technician	R 145	177	.....		24	29	.....	
	N 45	242	.....		7	40	.....	
Nurse	R 155	196	.....		25	32	.....	
	N 35	223	.....		6	37	.....	
SIGHT-SCREENER, by—								
Technician	R 144	138	.....		24	23	.....	
	N 46	281	.....		7	46	.....	
Nurse	R 138	191	.....		23	31	.....	
	N 52	228	.....		8	38	.....	
TELEBINOCULAR, study std., by—								
Technician	R 142	164	93	106	23	27	15	17
	N 48	255	44	363	8	42	8	60
Nurse	R 148	201	103	226	24	33	17	37
	N 42	218	34	243	7	36	6	40
TELEBINOCULAR, mfr's std., by—								
Technician	R 146	199	106	175	24	33	18	29
	N 44	220	31	294	7	36	5	48
Nurse	R 142	229	122	307	23	38	20	51
	N 48	190	15	162	8	31	3	26
TEACHER'S JUDGMENT and SNELLEN, high std., by—								
Teacher	R 125	145	98	153	21	24	16	25
	N 65	274	39	316	10	45	7	52
TEACHER'S JUDGMENT and MASSACHUSETTS by—								
Technician	R 141	170	102	182	23	28	17	30
	N 49	249	35	287	8	41	6	47
Nurse	R 143	169	103	183	23	28	17	30
	N 47	250	34	286	8	41	6	47
NEAR VISION, high std., by—								
Technician	R 69	43	72	51	11	7	12	8
	N 121	376	65	418	20	62	11	69
Nurse	R 80	69	81	124	13	11	13	20
	N 110	350	56	345	18	58	10	57

TABLE 2 a—Continued

Procedure and Tester		NUMBERS				PERCENT			
		Sixth grade (total = 609)		First grade (total = 606)		Sixth grade (total = 100)		First grade (total = 100)	
CRITERION—OPHTHALMOLOGIST'S JUDGMENT from CLINICAL EXAMINATION		R	N	R	N	R	N	R	N
		190	419	137	469	31	69	23	77
NEAR VISION, low std., by—									
Technician	R	28	5	21	3	5	1	4	1
	N	162	144	116	466	26	68	19	76
Nurse	R	40	13	38	24	7	2	6	4
	N	150	406	99	445	24	67	17	73
SNELLEN, high std., and NEAR VISION, high std., by—									
Technician	R	109	48	87	76	18	8	14	13
	N	81	371	50	393	13	61	9	64
Snellen by Teacher, Near Vision by nurse	R	120	89	98	144	20	15	16	24
	N	70	330	39	325	11	54	7	53
SNELLEN, high std., and NEAR VISION, low std., by—									
Technician	R	87	15	64	42	14	3	11	7
	N	103	404	73	427	17	66	12	70
Snellen by teacher, Near Vision by nurse	R	101	44	90	64	17	7	15	11
	N	89	375	47	405	14	62	8	67
TEACHER'S JUDGMENT, SNELLEN, high std., by teacher and NEAR VISION, high std., by nurse	R	142	172	111	213	23	28	18	35
	N	48	247	26	256	8	41	5	42

tests A, B, and C, respectively. The 3 scatters are reproduced below from table 2a; it will be recalled that the marginal totals taken vertically are the 190 referrals and 419 nonreferrals by the ophthalmologist.

Test A		R	N
Total	609	190	419
R	322	145	177
N	287	45	242

Test B		R	N
Total	609	190	419
R	99	85	14
N	510	105	405

Test C		R	N
Total	609	190	419
R	112	69	43
N	497	121	376

Two factors determine the distributions of referrals and nonreferrals by these tests. One is the efficiency of the test as a method of measurement, which, in this study, means its agreement with the criterion, ophthalmologist's judgment. The other factor is the cutoff point, or

TABLE 2 b

Correlations and Total Referrals by Screening Procedures

Correlations are "point" coefficients computed from the 2 x 2 scatters in the left-hand part of table 2a. Total referrals are sums of the upper figures in the percentage scatters of table 2a; for example, the 32 percent representing total referrals by Teacher Judgment in sixth grade is the sum of 13 and 19 percent.

Procedure	Tester	CORRELATIONS		TOTAL REFERRALS (percent)	
		Sixth grade	First grade	Sixth grade	First grade
Judgment	Teacher	0. 14	0. 16	32	29
Snellen, high std	Technician	. 52	. 42	16	18
	Nurse	. 44	. 42	20	20
	Teacher	. 44	. 48	21	22
Snellen, low std	Technician	. 38	. 33	8	5
	Nurse	. 35	. 43	7	5
	Teacher	. 37	. 34	10	5
Massachusetts	Technician	. 48	. 42	31	31
	Nurse	. 42	. 40	30	31
Ortho-Rater	Technician	. 32	.....	53	.....
	Nurse	. 33	.....	57	.....
Sight-Screener	Technician	. 40	.....	57	.....
	Nurse	. 25	.....	54	.....
Telebinocular, study std	Technician	. 33	. 40	50	32
	Nurse	. 28	. 23	57	54
Telebinocular, mfr's std	Technician	. 27	. 34	57	47
	Nurse	. 19	. 22	61	71
Judgment and Snellen, high std.	Teacher	. 29	. 33	45	41
Judgment and Massachusetts	Teacher and technician	. 31	. 30	51	47
	Teacher and nurse	. 32	. 30	51	47
Near Vision, high std	Technician	. 31	. 43	18	20
	Nurse	. 28	. 29	24	33
Near Vision, low std	Technician	. 28	. 31	6	5
	Nurse	. 29	. 31	9	10
Snellen, high std. and Near Vision, high std.	Technician	. 49	. 45	26	27
	Teacher and nurse	. 41	. 35	35	40
Snellen, high std. and Near Vision, low std.	Technician	. 52	. 42	17	18
	Teacher and nurse	. 46	. 50	24	26
Judgment, Snellen high std. and Near Vision, high std.	Teacher and nurse	. 31	. 30	51	53

standard for referral, used with the screening procedure. (The ophthalmologist's standard for referral is also a factor, but is relatively unimportant here because it is constant for each test.)

One would not judge the relative efficiency of a yardstick and a tape-

measure as measuring devices by comparing the results obtained when using a cutoff point at 30 inches for the yardstick and at 20 inches for the tape-measure. With such a procedure the yardstick would always throw more of the objects measured into the short pile. So, with the 3 screening tests, the fact that test A refers about three times as many students as B or C (322 as compared with 99 and 112) indicates that A's standard for referral is higher, but gives no information as to its efficiency in measuring.

The figures in the scatter give the essential information regarding screening efficiency, but they are so affected by the differences in total referrals that they are not readily comparable from one test to another. Of special interest in this connection are the missed cases (students referred by the ophthalmologist but not by the test) and the overreferrals. For tests A, B, and C, these are as follows:

	Missed cases		Over-referrals	
	Number	Percent of 190	Number	Percent of total referrals by test
Test A	45	24	177	55
Test B	105	55	14	14
Test C	121	64	43	38

We could suspect from these values that B is the most efficient of the 3 tests but we would be left uncertain as to how much better B is than the others, and as to whether A and C have substantially different, or about the same, efficiency. Similar difficulties are encountered in attempts to use any other percentages for comparison of the scatters.

For accurate comparison of the efficiency of one test with that of another we need a single expression, or index, of the efficiency of each. This index must take into account all of the categories in the 2 x 2 scatter. There is no percentage relationship or combination of percentages which could accomplish this in a consistent manner for all scatters in table 2a. But use of the "point correlation coefficient" solves the problem in a way that is altogether reasonable for the purpose. (The computation of this coefficient and the possible use of tetrachoric or other coefficients are discussed in appendix D.)

The correlation states the degree of correspondence or agreement between the screening procedure results and the ophthalmologist's judgment. At the same time it is practically independent of the proportion of total referrals. It therefore expresses the efficiency of the procedure as a method of screening regardless of the standard for referral employed.

In practice, the correlations obtained with different standards for referral are likely not to be quite the same owing to sampling errors or unevenness in the distributions of

scores in limited samples of the population. Moreover, the nearer to one extreme or the other the cutoff point is placed, the more the correlation is affected by chance; i. e., if the proportion of total referrals is low the correlation is not as stable statistically as when the proportion of total referrals is high.

For a procedure that consists of a battery of tests, changing the cutoff points may also affect correlation with the criterion by changing the relative weights of the subtests. But any material change in the relative weights of subtests would be equivalent to a change in the procedure itself.

The correlation coefficients for the 3 tests we have been considering are 0.32 for A, 0.52 for B, and 0.31 for C. From these values it is clear not only that B is substantially more efficient than A or C, but also that there is little, if any, difference in the efficiency of A and C.

Table 2b shows the correlation with clinical judgment of the results of each screening procedure, and the total percentage of referrals by each. These values summarize the most essential information to be derived from table 2a.

Relative efficiency of the procedures. We may now examine the evidence as to the relative efficiency of the different procedures, beginning with the correlations for sixth-grade students in table 2b. More correlations are available for some of the procedures than others. For the Snellen there are 6 (2 standards, with 3 testers each), for the Telebinocular 4 (2 standards, with 2 testers each), and for most of the others 2 each. We will take first the average of the available correlations for each procedure, disregarding for the moment possible differences among the testers or differences that might be related to cut-off points or standards.

For the Snellen Test the average correlation is 0.42, for the Massachusetts 0.45, for both the Ortho-Rater and Sight-Screener 0.33, for the Telebinocular 0.27, and for the Near Vision Test 0.29.

There is no rigorous procedure for testing the statistical significance of differences among the correlation coefficients. Nevertheless, if it is recognized that small differences could often arise by chance, the correlations yield valuable evidence concerning the relative efficiency of the procedures.

There is some evidence, which will be discussed later, that the technician was a slightly more efficient tester than the nurse. Consequently, where there is a difference between the correlations obtained by technician and nurse for a given screening procedure, the efficiency of that procedure, or the efficiency of which it is capable, is probably represented better by the correlation found for the technician than by the value found for the nurse.

With respect to the correlations for the high and low standard Snellen Test, it should be noted that only a small proportion of students are referred when the low standard is used. This condition, as mentioned earlier, means that the correlation is subject to a relatively large sampling error. In this light the value 0.47, which is the average of the 3 correlations obtained with the high standard, probably represents the overall

screening efficiency of the Snellen Test more accurately than the value 0.42, which was noted above as the average of the 6 correlations available for the high and low standards taken together. We may, therefore, believe that the Snellen's efficiency is best represented by a coefficient of about 0.45.

With respect to the Telebinocular, the study standard probably amounts to being a somewhat different procedure from the manufacturer's standard, insofar as the 2 standards do not give the same relative weights to the component subtests. It is thus likely that the difference between the average correlations available for each, which are 0.30 for the study standard and 0.23 for the manufacturer's standard, is not due entirely to chance variation.

A similar effect is found when the standard is changed for only one part of the Massachusetts Vision Test. From table 2b it is seen that the average of the correlations for this procedure as given by technician and nurse to first-grade students is 0.41. As indicated elsewhere the first part of the procedure is the Snellen Test with the high standard. We may ask what the scatters and correlations would be if the low standard Snellen were used instead, with no change in the other parts of the procedure. By tabulating the data for first-grade students accordingly, the following results are found:

Massachusetts Vision Test (with low standard for part I)		Total	Ophthalmologist		Correlation
			R	N	
		606	137	469	
By technician	R	128	56	72	0.26
	N	478	81	397	
By nurse	R	113	59	54	.34
	N	493	78	415	

Since the average correlation is only 0.30, it is clear that changing the standard of only one part of the procedure considerably changes the weight given to that part, and thus alters the procedure as a whole, substantially reducing its screening efficiency.

Teacher Judgment, as used in this study, has a very low correlation with the ophthalmologist's judgment—only 0.14. It is not surprising, therefore, that when it is combined with another procedure, the combination shows poorer correlation with the criterion than that for the other procedure alone. Thus Teacher Judgment with the high standard Snellen gives a correlation of 0.29, whereas 0.44 is found for the high standard Snellen alone, with teacher as tester; Teacher Judgment combined with the Massachusetts Vision Test correlates 0.31, as compared with the value 0.45 for the Massachusetts alone; the combination Teacher

Judgment, Snellen, and Near Vision correlates 0.31, while that for the Snellen and Near Vision Test combination, without Teacher Judgment, is about 0.45.

The combination of the Near Vision Test with the Snellen Test gives approximately the same correlations as the Snellen Test alone.

In brief, the levels of efficiency of the various screening procedures, as indicated by their correlations with ophthalmologist's judgment, may be summarized thus for the sixth grade:

1. Since none of the procedures or combinations of procedures under consideration show very marked correlation with the criterion, none can be said to possess high screening efficiency.
2. Inasmuch as the Massachusetts and the Snellen Test correlate approximately 0.45 with the criterion, these are the most efficient—or least inefficient—of the procedures tested.
3. Next in order are the Ortho-Rater and Sight Screener with correlations of 0.33, but it is doubtful whether either practical or statistical significance attaches to the margin which these procedures have over the Telebinocular and Near Vision tests, for which the correlations are of the order 0.27 and 0.29.
4. No gain, but rather a substantial loss in efficiency, is found when Teacher Judgment, as obtained in this study, is used in combination with another procedure. It is possible that a better method of obtaining Teacher Judgment, with more extended observation, and referral only after teacher-nurse conference, would make a greater contribution than was obtained in the study.
5. No marked improvement in screening efficiency is found for the combination of Near Vision Test with the Snellen over that for the Snellen alone.

With only minor differences, the above pattern of findings is evident also in the correlations obtained for the first-grade students. Here, too, the Snellen and Massachusetts are the most satisfactory of the procedures employed, since these 2 procedures each have an average correlation of approximately 0.41 with the criterion.

The average correlation is 0.34 for the Near Vision test and only slightly less, 0.30, for the Telebinocular. The Ortho-Rater and Sight-Screener were not administered in the first grade.

The correlation for Teacher Judgment, 0.16, is about the same as for sixth grade, and combining Teacher Judgment with other procedures has the same effect of lowering the correlation obtained for the other procedures alone. The correlations for the combinations of Snellen and Near Vision Test are little, if any, better than those for the Snellen alone.

Relative Efficiency of Testers. The correlations available separately for technician, nurse, and teacher are of interest also.

From the averages of the 4 correlations available for teacher as compared

with the other testers (high and low standards for both grades), it appears that, in this study, the teachers were able to administer the Snellen Test about as efficiently as the technician and nurses.

If each of the 3 procedures (Snellen, Telebinocular, and Near Vision) for which 2 standards were used is counted as 2 procedures, there are 9 procedures for the sixth grade and 7 for the first grade in which separate correlations are available for technician and nurse. Although the differences among these pairs of coefficients are not consistent as between one procedure or grade and another, an average difference of approximately 0.05 correlation points is found in favor of the technician. In view of his special training, this average difference is in the expected direction, and from a statisticial standpoint the difference is probably significant. Yet, since the absolute size of the difference is not very great, the correlations may be said to indicate that the screening efficiency of the nurse is not far below that of the technician.

Where the correlation obtained by the technician is markedly higher than that for the nurse, it is quite possible that some characteristic of the screening procedure makes it more difficult for the less experienced tester to administer it efficiently. Examples of such likely differences are seen in the correlations for the Telebinocular, first grade, and for the Sight Screener.

Variation in total referrals. While the efficiency of a screening procedure is measured by its correlation with the criterion, in the practical use of a procedure we are also concerned with the proportion of children referred. For any procedure, raising the standard of referral will increase the proportion of total referrals. And in the absence of perfect correlation with the criterion, which is not attainable in practice, the ratio of correct referrals to incorrect referrals will worsen as the proportion of total referrals rises.

If there were a high degree of correlation between the procedure and the criterion, this decrease in the ratio of correct to incorrect referrals might be of small consequence. Where the correlations are as low as those with which we are here concerned, there will be a point beyond which it is not practicable to raise the standards because the increase in correct referrals no longer compensates for the associated increase in incorrect referrals.

If we look now at the total referrals by the different procedures shown in table 2b we see, as expected, a higher proportion of total referrals by the Snellen and Near Vision Tests when the high standards are applied than with the low standards. But even the high standard Snellen and Near Vision Tests give fewer total referrals than any of the other procedures.

It is not surprising that the other procedures give more referrals when one considers their construction. A Snellen Test consists of only 2 measures, 1 of each eye. In the Near Vision Test there are only 3 measures. But the Massachusetts Vision Test has 7 measures, and there

are 11 to 14 each in the Ortho-Rater, Sight-Screener, and Telebinocular Tests.

Adding a measure to a procedure tends to have an effect like that of raising the standard for referral. It may also change the overall correlation with the criterion, and for that reason the effect is not identical with that of simply raising the standard for 1 measure or a group of similar measures like those in the Snellen or Near Vision Tests. Yet the effects are similar to the extent that the requirements for "passing" are increased in both cases, and, if the cutoff points for measures already included are unchanged, the addition of a new measure necessarily increases the proportion of referrals.

No constant relation is to be expected between the number of measures in a procedure and the proportion of referrals, if only because this relation

TABLE 3

Effect of Retesting, With Referral of Students Failing Both Times

A referral (R) by 2 administrations of a screening procedure is a student who was referred twice, i. e., by both technician and nurse; a nonreferral (N) by two administrations is a student who was referred only once, or by neither tester. Each value in the last column is the average of the correlations shown in table 2b for the corresponding procedure, as obtained independently by technician and nurse.

Grade and procedure	Ophthalmologist's judgment		Correlation for the two administrations of the screening procedure	Average correlation for single administration
	R	N		
Sixth Grade:				
Totals	**190**	**419**		
Massachusetts Vision Test R	96	30	0.50	**0.45**
Massachusetts Vision Test N	94	389		
Telebinocular, Study Standard R	128	112	.39	**.31**
Telebinocular, Study Standard N	62	307		
First Grade:				
Totals	**137**	**469**		
Massachusetts Vision Test R	81	38	.54	**.41**
Massachusetts Vision Test N	56	431		
Telebinocular, Study Standard R	81	67	.44	**.32**
Telebinocular, Study Standard N	56	402		

is affected by the cutoff points applied and the intercorrelations of the measures. Teacher Judgment, for example, is a relatively independent measure and its proportion of referrals is higher than that for most of the individual subtests. All of the combinations in which Teacher Judgment is included refer 45 to 51 percent of the sixth-grade students, although the number of subtests varies from 3 for Judgment and Snellen, to 8 for Judgment and Massachusetts Vision Test.

Except as regards a measure like Teacher Judgment, however, some correspondence between the number of measures or subtests, and the proportion of total referrals is to be expected where the cutoff points for individual measurements are at fairly comparable levels. This was true for most of the procedures used in this study, so that from one procedure or combination of procedures to another a direct, if rough, relation appears between the number of component subtests and the proportion of referrals.

The high standard Snellen and Near Vision Tests, with 2 or 3 measures each, refer about 20 percent of the sixth-grade students. The combination of these tests, comprising 5 measures, refers approximately 30 percent, and so also does the Massachusetts, which has 7 measures. The proportions referred are much higher, ranging between 50 and 60 percent, for the Ortho-Rater, Sight-Screener and Telebinocular, each of which has 11 to 14 measures.

With any multiple-test procedure, it is of course possible to reduce total referrals by lowering the cutoff points of the component tests. If the cutoffs are lowered more or less proportionately for all components, the procedure's efficiency, or correspondence with the criterion, is not likely to be affected. The effects of change are nevertheless difficult to predict. If the change occasions a shift in the relative weights of the subtests, it may result either in greater efficiency—as in the case of the study standard Telebinocular—or in poorer correspondence with the criterion—as when a lower standard is used for part I of the Massachusetts Vision Test.

Effect of Retesting Before Referral

In screening programs it is frequently the practice to administer the screening procedure a second time to those students who fail a first test, and to refer only those who fail both times. To see the effect of retesting in this way, the data for the Massachusetts Vision Test and the Telebinocular, study standard, have been tabulated to show how many students were referred by *both* testers. The results are given in table 3.

The extent to which screening efficiency is improved by this use of retests is easily misjudged unless, as in this study, all of the students are examined by an eye specialist. A reduction in over-referrals is more impressive in the absence of evidence as to the associated reduction in correct referrals than it is when the full data can be examined. In

table 3, over-referrals of sixth grade students by two administrations of the Massachusetts Vision Test are only 30, as compared with 65 or 72 by technician alone or nurse alone (table 2a). Yet the repetition has also reduced the number of correct referrals to 96, instead of the 120 or 112 obtained by technician or nurse alone.

Comparison of the correlation coefficients gives the best evidence of the degree of improvement in screening efficiency. All 4 of the correlations shown for 2 administrations of a procedure are higher than those for single administrations, indicating better correspondence with the ophthalmologist's judgment. Yet the improvement is not very great.

Repeating a screening procedure and referring only those students who fail both times gives enough gain in screening efficiency to be worth while if other methods of improvement are not available or administratively feasible. But it is clear that a single repetition of a procedure as a whole cannot be relied upon as an effective remedy for low screening efficiency.

Reliability and Validity of Component Parts of Screening Procedures

The efficiency of a screening procedure depends upon the reliability and validity of its component parts. The study has provided data on the reliability and validity of the component parts of the various procedures when used for testing sixth- and first-grade students. These data are largely in terms of correlation coefficients. Unlike the "point" correlations used in the other sections, the correlations here are coefficients computed in the usual way from scatters having several class intervals for each variable.

Reliability Coefficients. As obtained in this study the reliability coefficient is the test-retest correlation, or the correlation showing the extent to which 2 testers agree with themselves or with each other when repeating the same test on the same children.

Table 4 shows the test-retest correlations obtained when the children were retested by a second tester. Except where otherwise noted the correlations are based on tests and retests of 609 sixth-grade students, or 539 white first-grade students. A random half of the children in each grade were tested first by the nurse, the other half were tested first by the technician.

Table 5 shows test-retest correlations when the retest was given by the same tester who administered the first test. Fifty-seven sixth-grade students and 55 first-grade students were retested by the same tester.

Discussion of the test-retest correlations will be based chiefly on table 4 because reliability coefficients are more meaningful where first and second tests are given by different persons, and because there are more

TABLE 4

Test-Retest Correlations: Technician vs. Nurse

	Sixth grade					First grade [1]		
	O.–R.	S.–S.	Tel.	M.V.T.	N.V.	Tel.	M.V.T.	N.V.
Acuity, far, right	0.65	0.74	0.75	0.72	[2]	0.52	0.61	[2]
Acuity, far, left	.66	.76	.77	.70	[2]	.59	.64	[2]
Acuity, far, both	.64	.74	[2]	[2]	[2]	[2]	[2]	[2]
Acuity, near, right	.58	.74	.53	[2]	.74	.45	[2]	.49
Acuity, near, left	.69	.74	.60	[2]	.72	.46	[2]	.45
Acuity, near, both	.57	.65	.49	[2]	.74	.51	[2]	.53
Phoria, far, lateral	.71	.77	.72	.54	[2]	.65	.27	[2]
Phoria, near, lateral	.69	.82	.48	.31	[2]	.54	.29	[2]
Phoria, far, vertical	.54	.21	.12	[3]	[2]	[3]	[3]	[2]
Phoria, near, vertical	.60	.24	[2]	[2]	[2]	[2]	[2]	[2]
Stereopsis, far	.60	.46	.74	[2]	[2]	[4].47	[2]	[2]
Stereopsis, near	[2]	.64	[2]	[2]	[2]	[2]	[2]	[2]
Fusion, far	[2]	[2]	.45	[2]	[2]	.40	[2]	[2]
Fusion, near	[2]	[2]	.37	[2]	[2]	.35	[2]	[2]
Plus sphere, right	[2]	[2]	[2]	.62	[2]	[2]	.39	[2]
Plus sphere, left	[2]	[2]	[2]	.57	[2]	[2]	.47	[2]
Binocular vision, far	[2]	[5].63	[2]	[2]	[2]	[2]	[2]	[2]
Binocular vision, near	[2]	[5].70	[2]	[2]	[2]	[2]	[2]	[2]

[1] White students only except where noted otherwise.

[2] Test not part of the screening procedure, or not used in deciding referral.

[3] Too little variation from modal scores for meaningful correlation; all but a few cases (less than 10) fell at mode of one or both variables.

[4] Negro students only.

[5] In computing these coefficients scores of "right only," "left only" and "alternating" were grouped into one category and placed between the categories "binocular vision" and "reads none."

than 10 times as many retests by the second tester as by the same tester.

Considering first the correlations in table 4 for sixth-grade students, we find that the coefficients for tests of visual acuity at far point by the Sight-Screener and Telebinocular are approximately 0.75; those for the Massachusetts Vision Test are nearly as high, while those for the Ortho-Rater are somewhat lower.

Average correlations for the Sight-Screener test of visual acuity at

near point and the Near Vision Test are a little over 0.70, but for the parallel test by the Ortho-Rater the average is nearer 0.60, and it is even lower for the Telebinocular's corresponding tests.

Among the tests for lateral heterophoria at far point, the Massachusetts has a relatively low coefficient (0.54), but the values are over 0.70 for the other 3 tests. For the tests of lateral heterophoria at near point the variation is greater, ranging from 0.31 for the Massachusetts to 0.82 for the Sight-Screener.

In the Sight-Screener tests for lateral heterophoria the student is asked to report where he sees the arrow when he first looks at the test target, and then to tell where he sees the arrow after it has "stopped moving." The responses are recorded as the "first reading" and "second reading." The correlations for the Sight-Screener lateral heterophoria tests shown in the tables are for the second reading. Retest correlations (technician vs. nurse) for the first reading are: far point, 0.61: near point, 0.67. It seems probable that the higher correlations shown in table 4 for the Sight-Screener tests for lateral heterophoria as compared with those for other tests for lateral imbalance may be due more to the method of administration of the test than to its construction.

The reliability of the tests of vertical heterophoria is generally unsatisfactory. That for the Ortho-Rater would seem to show some promise except for the fact that its validity is uncertain, as will be noted below.

As shown in tables 25–28 of the Appendix, the scores on measures of vertical heterophoria are heavily concentrated at the modal score, which merely means that all but a small percentage of students are found to be orthophoric. This is not undesirable as far as screening work is concerned, but it does mean that the correlations for vertical heterophoria tend to be quite unstable, and that caution is needed in interpreting them. The test-retest coefficient of 0.31 shown in table 4 for the Massachusetts test of near lateral phoria is also subject to this difficulty. The coefficient is based on over 600 cases and its standard error would be 0.04 by the usual formula. However, the scatter for this correlation showed that only 11 cases fell outside the modal score for technician, nurse, or both testers, and with another sample of 600 children the correlation could easily be 0.20 or 0.60 depending on how a dozen or so cases fell. This example has been used because it concerns the least stable of the coefficients shown. Some scatters had even less meaning statistically, and, as indicated in the table, the correlations have been computed only for scatters where 10 or more students made scores differing perceptibly from orthophoria in both tests.

The correlations of the Telebinocular tests of fusion are about 0.40. Those for the various tests of stereopsis are higher, particularly for the Telebinocular. The correlations for stereopsis by the Sight-Screener are based on scores for the best reading the student made without previous error. Scoring for the best reading made by the student regardless of

previous error gave lower correlations (0.44 and 0.52, respectively, for far and near stereopsis).

The plus sphere test of the Massachusetts Vision Test yielded moderately good reliability, the coefficients for both right and left being close to 0.60. The values for the Sight-Screener tests of binocular vision are also substantial, though they are unstable because at far point only 15 students, and at near point only 13 students, were not found to have binocular vision by technician, nurse, or both.

Many of the "same tester" coefficients in table 5 tend to be higher than

TABLE 5

Test-Retest Correlations: Same Tester

	Sixth grade					First grade	
	O.-R.	S.-S.	Tel.	M.V.T.	N. V.	Tel.	M.V.T.
Acuity, far, right	0.72	0.58	0.81	0.80	[2]	0.61	0.63
Acuity, far, left	.79	.89	.79	.68	[2]	.84	.38
Acuity, far, both	.78	.75	[2]	[2]	[2]	[2]	[2]
Acuity, near, right	.35	.75	.52	[2]	[1].00	.66	[2]
Acuity, near, left	.77	.76	.53	[2]	.85	.66	[2]
Acuity, near, both	.56	.73	.66	[2]	[1].97	.77	[2]
Phoria, far, lateral	.75	.78	.68	.55	[2]	.86	.41
Phoria, near, lateral	.71	.78	.63	.80	[2]	.42	.71
Phoria, far, vertical	.51	.49	[1].72	.00	[2]	.55	.00
Phoria, near, vertical	.70	.76	[2]	[2]	[2]	[2]	[2]
Stereopsis, far	.79	.84	[1].99	[2]	[2]	.66	[2]
Stereopsis, near	[2]	.74	[2]	[2]	[2]	[2]	[2]
Fusion, far	[2]	[2]	.71	[2]	[2]	.63	[2]
Fusion, near	[2]	[2]	.70	[2]	[2]	.31	[2]
Plus sphere, right	[2]	[2]	[2]	.74	[2]	[2]	.51
Plus sphere, left	[2]	[2]	[2]	.83	[2]	[2]	.69
Average	.68	.74	.67	.63	.85	.63	.48
Average for nurse only (33 students)	.54	.52	.67	.65 (.60=Nurse's over-all av.)			
Average for technician only (24 students)	.82	.74	.75	.57 (.72=Technician's over-all av.)			

Coefficients for sixth grade are based on 57 students and coefficients for first grade are based on 55 students except for Sight-Screener phoria, near lateral, which is based on 24 students tested by the technician.

In each grade approximately half the students were tested and retested by the technician and half were tested and retested by the nurse.

[1] Excluded in computing averages because the scatters for these coefficients showed too little variation from modal scores for meaningful correlations.

[2] Test not a part of the screening procedure.

the coefficients for the same tests when repeated by a different tester. It may be that on some of these tests the individual testers make certain errors that they tend to repeat on retesting, so that the errors are correlated in the first and second tests, with higher test-retest coefficients for "same tester" as a result. A simpler and quite as likely explanation is that the smaller samples of children on whom the table 5 coefficients are based did not fully represent the total groups.

TABLE 6

Validity Coefficients: Correlations Between Parts of Screening Procedures by Technician and Parallel Parts of Clinical Examination

	Clinic Test	Sixth grade				First grade [1]	
		O.-R.	S.-S.	Tel.	M.V.T.	Tel.	M.V.T.
Acuity:							
Far, right	Snellen	0.62	0.68	0.67	0.71	0.51	0.54
Far, left	Snellen	.62	.67	.69	.71	.54	.54
Heterophoria:							
Far, lateral	Rod	.40	.61	.49	.54	.33	[2]
Far, lateral	Cover	.24	.33	.34	.44	.40	[2]
Near, lateral	Rod	.33	.59	.36	.42	.27	[2]
Near, lateral	Cover	.28	.47	.31	.45	.43	[2]
Near, lateral	Wing	.27	.56	.33	.35	.16	[2]
Far, vertical	Rod	.16	−.10	[2]	[2]	[2]	[2]
Far, vertical	Cover	[2]	[2]	[2]	[2]	[2]	[2]
Near, vertical	Rod	.07	−.14	[3]	[3]	[3]	[3]
Near, vertical	Cover	[2]	[2]	[3]	[3]	[3]	[3]
Near, vertical	Wing	−.01	−.08	[3]	[3]	[3]	[3]

In computing each of these coefficients, students on whom for any reason a measurement was not obtained by either the clinician or the technician were excluded. No sound method of including these cases was found; at the same time it was not apparent that their exclusion tended to bias the correlations. Some idea of the numbers of cases excluded may be obtained, if desired, from the "no reading" frequencies shown in the tables of Appendix C, or, specifically, from the sum of these frequencies for the two measures concerned in each correlation. This sum gives, however, the maximum possible exclusions (and therefore tends to overstate the actual exclusions) from each scatter, since the students actually excluded are to some extent the same for both the screening test and the clinical test.

[1] White students only.

[2] Too little variation from modal scores for meaningful correlation; all but a few cases (less than 10) fell at mode of one or both variables.

[3] Test not part of the screening procedure.

The reliability coefficients for first-grade students in table 4 are generally low. The coefficients for far acuity and the Telebinocular tests of lateral heterophoria are relatively good, but in the main the correlations are not high enough to expect very effective screening from the tests as they stand.

Validity Coefficients. For the purpose of this study, validity coefficients, which are shown in table 6, may be defined as the correlations between parts of the screening procedures and parallel parts of the clinical examination. The measurements by the clinical examination are taken as criteria of the effectiveness of the tests in the screening procedures, not because the former are necessarily nearer the true values, but because the usefulness of a screening procedure depends upon the degree of correlation with the clinical examination.

In general, the magnitudes of validity coefficients are limited by the magnitudes of the reliability coefficients. That is to say, if the measurements of any one variable, when repeated, are not consistent enough to correlate with themselves, they cannot be expected to correlate with any other variable.

On the assumption that the clinical tests have about the same reliability as the screening tests, it would be possible, theoretically, to use the available data on reliability to correct the validity coefficients for the effects of errors of measurement. In this report, no table of validity coefficients "corrected" in this manner is offered because the data do not meet all the conditions requisite to such adjustments. In view of the relatively low test-retest reliability of many of the screening tests, it is obvious that there are many chance errors in the measurements and that without these the correspondence between the screening tests and the ophthalmologist's tests would be considerably higher than the coefficients shown in table 6.

When the test-retest coefficients are taken into account, one can conclude from the validity coefficients that most of the component parts of the screening tests would have high enough validity for screening purposes if the scores obtained were highly reliable. But the uncorrected validity coefficients show the degree of correlation between the clinical tests and the screening tests as they are actually performed with present methods of administering the tests. In this light the tests of visual acuity at far point show a high enough correlation with the parallel clinical test to be used for screening purposes, and this may be true also for selected tests of lateral heterophoria, particularly with the Sight-Screener. The coefficients of the other heterophoria tests are not reliably high enough to indicate efficient screening.

The correlations for vertical heterophoria are especially unstable due to the high concentrations of scores at modal values. The validity coefficients shown in table 6 for far and near vertical heterophoria range from -0.14 to 0.16, suggesting that all values for these tests would be close to zero if very large groups of students were used. It is nevertheless

conceivable that one or two of these tests would have shown satisfactory validity if, say, a thousand more students had been tested.

Another way of examining the data is to see to what extent the clinician and the screening tests obtained measurements indicative of vertical heterophoria on the same students. For this purpose it is difficult to interpret the cases on which no reading is obtained, so only those students are considered hyperphoric who, according to the clinical or screening test in question, had a measurement of over 1 prism diopter of vertical imbalance.

None of the technician's screening tests gave measurements indicative of hyperphoria at far point on any of the 11 sixth-grade students who were hyperphoric according to the clinician's Maddox rod test at 20 feet. Of the 22 sixth-grade students that the Maddox rod test showed to have hyperphoria at near point, the Sight-Screener near vertical phoria test referred none. The parallel test by the Ortho-Rater referred 3, but on two of these the findings of the two tests were contradictory as to the direction of the deviation.

In summary, as far as the evidence in this study carries, there is no indication that the available tests of vertical heterphoria have any substantial value for screening work, at least not as the tests stand.

At the first-grade level the Massachusetts tests of lateral heterophoria are subject to the same difficulty. For both the Telebinocular and the Massachusetts, only the coefficients for far acuity are substantial enough to suggest efficiency for screening first-grade students.

No tests of visual acuity at near point are included in the clinical examination, but the intercorrelations between the different screening tests of this function are of interest as an indication of the extent to which they measure the same thing (table 7). They range from 0.55 to 0.69 for the technician's tests, and from 0.48 to 0.62 for the nurse's tests of sixth-grade students. Since the 12 sixth-grade test-retest coefficients of these 4 tests average only 0.65 (table 4), it appears that the correspondence between the different tests is nearly as high as the correspondence between 2 tests with the same procedure by different testers.

For evaluation of the special tests of stereopsis and fusion in the screening procedures we have only the test-retest reliability coefficients (table 4) and the distributions of measurements according to Clinical Judgment presented in Appendix C, tables 29, 30, and 31. These distributions show that there are definite, if moderate, associations between these parts of the screening procedures and the ophthalmologist's judgment based on his complete examination, and that the associations are higher for sixth grade than for first grade. But the test-retest reliability coefficients are low for all but the Telebinocular Test of stereopsis.

The second part of the Massachusetts Vision Test, which uses plus sphere lenses as a test for latent hyperopia, has no parallel test in the clinical examination. The clinical measurement with which it might be expected to show the closest association is the spherical equivalent. Table 8

TABLE 7

Intercorrelations of Different Tests of Visual Acuity at Near Point (Averages of Correlations for Right Eye and Left Eye)

	Technician's tests					Nurse's tests				
	Sixth-grade students									
	Near Vision	Ortho-Rater	Sight-Screener	Tele-binocular	Average	Near Vision	Ortho-Rater	Sight-Screener	Tele-binocular	Average
Near Vision	.....	0.57	0.69	0.55	0.60	.....	0.58	0.62	0.48	0.56
Ortho-Rater	0.57	.....	.66	.62	.62	0.58	.....	.61	.58	.59
Sight-Screener	.69	.66	.....	.59	.65	.62	.61	.....	.56	.60
Telebinocular	.55	.62	.59	.....	.59	.48	.58	.56	.....	.54
Average	.....	.....	.....	.....	.62	.....	.....	.....	.....	.57
	First-grade white students									
Telebinocular	0.56	.....	.....	.....	.....	0.40	.....	.....	.....	.....

TABLE 8

Distribution of Scores on M. V. T. Plus Sphere Test According to Spherical Equivalent Measurements

Plus sphere scores by technician	Total eyes	Spherical equivalent measurements in diopters on right and left eyes					
		Zero or minus values	+0.25–0.75	+1.00–1.75	+2.00–2.75	+3.00 or more	
Sixth-grade students							
Total	1,218	144	564	396	56	58	
Reads 20/20	54	0	5	15	21	13	
Reads 20/30	110	1	30	47	13	19	
Cannot read	1,054	143	529	334	22	26	
First-grade students						+3.00–3.75	+4.00 or more
Total	1,078	43	261	574	122	41	37
Reads 20/20	59	0	5	23	12	13	6
Reads 20/30	144	2	21	71	33	9	8
Cannot read	875	41	235	480	77	19	23

shows the distributions of measurements on the plus sphere test according to the measurements for spherical equivalent on the same eyes.

It is evident that there is a substantial association between the scores obtained on the plus sphere test and the degree of hyperopia or hyperopic astigmatism shown by the spherical equivalents. But it is also evident that the majority of eyes which, with the plus sphere lenses, were able to read one or both of the test lines, had spherical equivalents of less than +3.00 diopters.

Contributions of Specific Parts of Tests to Total Tests

The data that have been presented on test-retest reliability and validity of the component parts of the screening procedures give the most essential information as to the contribution the various subtests are capable of making to the total test, but they do not show how many students are

TABLE 9

Students Who Failed Component Parts of Massachusetts Vision Test Who Did Not Fail Previous Parts of the Test

Example: The first line concerns the 120 students referred by the technician's Massachusetts Vision Test out of the 190 sixth-grade students referred by the ophthalmologist. Of these, 85 failed the Massachusetts subtest of acuity; of those who passed that test, 23 failed the plus sphere subtest; of those who passed both those tests, 5 failed the far vertical phoria test; etc.

Grade	Tester	Ophthalmologist's judgment	Total M. V. T. referrals	Acuity	Plus sphere	Phoria		
						Far vert.	Far lat.	Near lat.
6	Technician	Refer.....(190 students)	120	85	23	5	4	3
		Nonrefer....(419 students)	65	14	34	3	8	6
6	Nurse	Refer.....(190 students)	112	87	14	2	6	3
		Nonrefer....(419 students)	72	34	26	2	6	4
1	Technician	Refer.....(137 students)	92	64	21	5	0	2
		Nonrefer....(469 students)	96	41	47	3	2	3
1	Nurse	Refer.....(137 students)	89	70	16	2	1	0
		Nonrefer....(469 students)	99	51	44	1	2	1

TABLE 10

Students Who Failed Component Parts of Telebinocular Test Who Did Not Not Fail Previously Listed Parts (Manufacturer's Standard)

Example: The first line concerns the 146 students referred by the technician's Telebinocular Test out of the 190 sixth-grade students referred by the ophthalmologist. Of these, 60 failed the Telebinocular subtest of far acuity; of those who passed that test, 8 failed the far vertical phoria test; of those who passed both of those subtests, 33 failed the test for lateral phoria at far point; etc.

Grade	Tester	Ophthalmologist's judgment	Total telebinocular referrals	Far point			Near point		Fusion far or near	Stereopsis
				Acuity	Phoria		Acuity	Phoria Lat.		
					Vert.	Lat.				
6	Technician	Refer.....(190 students).....	146	60	8	33	5	37	3	0
		Nonrefer....(419 students).....	199	22	3	66	0	104	3	1
6	Nurse	Refer.....(190 students).....	142	84	6	31	1	18	0	2
		Nonrefer....(419 students).....	229	27	6	81	3	105	5	2
1	Technician	Refer.....(137 students).....	106	74	2	13	5	6	0	6
		Nonrefer....(469 students).....	175	55	2	33	7	29	8	41
1	Nurse	Refer.....(137 students).....	122	83	3	18	10	7	0	1
		Nonrefer....(469 students).....	307	133	5	65	45	28	3	28

referred by one part of a procedure who are not referred by other parts. Examination of the data from the latter point of view has limited value because the number of students who fail one part of a procedure but not others will vary with the pass-fail cutoff points used for each part, as well as with the sequence in which the subtests are considered. It may be of interest, however, to look at some of the study data in this way.

In table 9 the Massachusetts Vision Test findings are presented from this point of view, showing the additional referrals by successive subtests when these are administered in the usual sequence. On the basis of the pass-fail standards used in the study the table shows for each subtest the number of students who failed that subtest but who did not fail any part of the procedure for which data are presented in a column to the left of that for the subtest in question.

Table 10 is a similar presentation of the results of the Telebinocular Test with the manufacturer's standard for referral. The sequence in which the component parts of the procedure are arranged has been selected as that which probably gives the most useful information about the contribution of each part of the procedure to the overall results.

Factors That Might Influence Test Results

Several factors that might influence the results of screening tests deserve consideration either because of their possible effect on the results obtained in this study or because of their bearing on the administration of screening programs. Some of these factors relate to the interpretation of test findings or the administration of the tests, others to the children who were the test subjects in this study.

Tests of Acuity Without and With Occlusion of Opposite Eye. The tests of visual acuity with the Ortho-Rater, Sight-Screener and Telebinocular are designed to be administered without occlusion of the opposite eye. Instructions for the Ortho-Rater tell the tester to repeat the tests of far and near acuity with occlusion whenever, without occlusion, the subject obtains a score of less than 20/22. The Telebinocular instructions call for repetition of the test of far acuity with occlusion if the score without occlusion is less than 20/25, and they state that if the score obtained with occlusion is better than that without occlusion this is evidence of suppression. These instructions were followed in administration of the Ortho-Rater and Telebinocular Tests in the study.

The instructions for the Sight-Screener Test supplied by the manufacturer at the beginning of the study did not call for any determinations of visual acuity with the opposite eye occluded. But at about the midpoint of the study a company representative requested that all tests of acuity be repeated with occlusion of the opposite eye. Accordingly, for about half of the sixth-grade students the tests of far and near acuity by the Sight-Screener were repeated with occlusion, regardless of the scores previously obtained without occlusion.

The scores obtained with occlusion were not taken into account in determination of whether or not a student was classed as a referral on the basis of his performance in the screening procedure.

Table 11 shows the numbers of eyes tested with occlusion by each procedure and by each tester, and the median scores obtained without and with occlusion. For both the Ortho-Rater and the Telebinocular the eyes tested include some that had scores without occlusion higher than the levels at which a test with occlusion was required. On sixth-grade students the Telebinocular tests by the nurse include 134 such eyes and the technician's only 26, which accounts for the higher median scores shown for the nurse. The group tested with occlusion by the Sight-Screener were not selected on the basis of scores obtained without occlusion, and both testers gave the test to the same students.

There is a slight but consistent tendency for the Ortho-Rater and Telebinocular scores obtained with occlusion to be higher than those obtained without occlusion. This does not appear in the Sight-Screener tests. If, instead of considering for the Sight-Screener all the eyes that were tested with occlusion, only those are considered which, without occlusion, had scores of less than 20/20, the tendency to higher median scores with occlusion is still absent.

The right-hand half of table 11 shows the numbers of eyes that without occlusion had scores that fell in the range requiring referral, but had satisfactory scores when retested with occlusion. It also shows what these numbers represent in percent of the total groups. The nurse consistently brought more students up to a "passing" score than did the technician, but she also obtained scores with occlusion that were lower than the scores without occlusion more frequently than did the technician.

It is unfortunate that no tests were given in reverse order, testing with occlusion first and retesting without occlusion. From the data available it is not possible to determine how much of the improvement in the scores is due to occlusion and how much is due to the learning factor.

Errors in Recording Test Results. Accuracy in recording test scores is primarily a matter of the efficiency of the tester, although it is related to the testing procedure to the extent that the simpler the test the fewer the opportunities for error.

By checking all records before the close of each day's testing session and keeping a record of all errors of recording noted it was possible to determine each tester's accuracy in recording test findings. For this purpose, only errors that were significant for interpretation of the test itself were considered. Errors on items such as the date or time of day were checked for purposes of the study but are not included in relation to this evaluation of efficiency in administering the tests. Most of the errors that could be identified were either omission of an entry for a part of a test, or more than one entry. A few incorrect entries could be recog-

TABLE 11

Tests of Acuity Administered Without and With Occlusion

Test of acuity	Technician			Nurse			Eyes for which scores were below "passing" without occlusion but satisfactory with occlusion			
	Eyes tested, number	Median scores [1]		Eyes tested, number	Median scores [1]		Number		Percent [2]	
		Without occl.	With occl.		Without occl.	With occl.	Tech.	Nurse	Tech.	Nurse
	Sixth grade									
Ortho-Rater:									*(Percent of 1,218)*	
Far	184	7.6	7.8	241	7.5	7.8	13	32	1	3
Near	149	8.3	8.6	200	8.4	8.9	22	32	2	3
Sight-Screener:									*(Percent of 628)*	
Far	628	7.5	7.5	628	7.2	7.2	13	35	2	6
Near	628	7.4	7.4	628	7.2	7.2	11	29	2	5
Telebinocular:									*(Percent of 1,218)*	
Far	170	3.1	3.8	288	4.7	6.1	22	41	2	3
	First grade									
Telebinocular:									*(Percent of 1,212)*	
Far	190	3.6	4.1	367	3.7	4.3	19	62	2	5

[1] Snellen equivalents of respective test scores are as follows: Ortho-Rater scores 7, 8, 9=20/30, 20/25, 20/22; Sight Screener scores 6, 7, 8=20/30, 20/20, 20/15; Telebinocular scores 3, 4, 5, 6=20/40, 20/28, 20/25, 20/22.

[2] For Ortho-Rater and Telebinocular 100 percent is the total number of eyes tested in sixth or first grade, since all those with low scores without occlusion were retested with occlusion. For Sight-Screener 100 percent is 628 eyes since only 314 students were tested after the test with occlusion was introduced.

nized, but as a rule it was not possible to know whether the tester had checked the score he intended to check. Presumably, however, this type of error might be expected with about the same frequency as omissions or multiple entries.

In the course of testing the 1,215 students, the technician was found to have made 26 errors on 23 records. Of these 18 were omissions, 4 were multiple entries, and 4 were incorrect entries that could be recognized. The nurses made 96 errors on 61 records, of which 73 were omissions, 10 multiple entries, and 13 incorrect entries.

It was observed that the nurses tended to make more errors when they first began to give the tests, and were more careful after several errors had been called to their attention. The routine checking of all records and the returning of them to the testers for correction before the testing session was completed doubtless made both testers especially conscious of the importance of accurate recording and tended to keep down the total number of errors.

The findings of the study in relation to accuracy of recording demonstrate again the importance of experience for efficiency in conducting a screening program.

Learning Factor. Since each student in the study was given each test twice, and the Snellen Test 3 times, it is of interest to know whether a learning factor affected the student's performance the second or third time he was given a test. To investigate this possibility, the scores obtained on certain subtests the first time the students were tested are compared with the scores obtained the second or third time they were given these tests. The scores examined are those for visual acuity of the right eye at far point and for lateral heterophoria at far point by each procedure, and the test of visual acuity at near point by the Telebinocular.

This comparison relates to the scores obtained when the entire testing procedure is repeated by the second tester after an interval of 1 to 24 hours, or occasionally longer. The learning factor in immediate repetition of a test might be more important.

TABLE 12

Distribution of Scores for Snellen Test of Right Eye of First-Grade Students the First, Second, and Third Times They Were Tested

Scores	First test	Second test	Third test
Total students	606	606	606
Poorer than 20/30	22	22	18
20/30	67	64	51
20/20 or better	517	520	537

TABLE 13

Median Scores[1] for Visual Acuity of Right Eye at Far Point by Telebinocular According to Whether the Tester Was the First or Second Person to Administer the Test to These Students

Students	Technician[2]		Nurse[2]	
	As first tester	As second tester	As first tester	As second tester
Sixth grade, white	9.3	9.5	9.0	9.3
Sixth grade, Negro	9.5	9.4	9.5	9.6
First grade, white	8.0	8.1	5.8	7.8
First grade, Negro	7.1	7.5	6.0	9.2

[1] Snellen equivalents of Telebinocular scores 5, 6, 7, 8, 9, and 10 are, respectively 20/25, 20/22, 20/20, 20/18, 20/17, 20/15.

[2] Each tester was first tester for approximately half of the 474 white and half of the 135 Negro sixth-grade students, and similarly for the 539 white and 67 Negro first-grade students.

For sixth-grade students no consistent difference could be recognized in the scores obtained on the first and later administrations of the tests of acuity at far point.

For first-grade students there was some evidence that a learning factor influenced the scores obtained for far acuity. Table 12 shows the distribution of first-grade students according to the scores they obtained on the Snellen Test the first, second and third times they were tested. The number of students scoring poorer than 20/20 is small on each testing and the significance of the differences in the distributions is doubtful, but there is some trend toward better performance with repeated testing.

With the Telebinocular test of far acuity (table 13) the technician obtained approximately the same median score on the group of first-grade students for whom he was the first tester as on the group who had been given the test previously by the nurse. But the nurse obtained much lower scores with this test on the first-grade students for whom she was first tester than on those whom the technician had already tested. This finding suggests that the technician was more successful than the nurses in helping the younger children understand what was expected of them and that as a result the students taking the test for the first time did as well for him as those previously tested by the nurse, while on the nurse's tests the students who had previously been instructed by the technician gave a better performance than those without this advantage.

For Telebinocular near acuity, the nurse obtained in both grades somewhat higher scores with students for whom she was second tester than with those for whom she was first tester. There was little difference in the scores obtained by the technician according to whether or not the students had been tested previously by the nurse.

No differences in the distributions of measurements for heterophoria appeared to be related to previous experience with the test.

A learning factor could affect the comparability of the overall findings of the study if the different screening tests had been given in the same sequence to all, or a majority, of the students. But since the testing sequence was varied, any effects on performance in one test from previous experience with other tests should be approximately the same for all the procedures studied.

Likewise, the fact that the technician and nurse were each first tester with equal frequency would not give either one an advantage over the other if the 2 testers were equally efficient. But the findings indicate that while previous testing by the nurse did not help the students to give better performance when tested by the technician, the nurse obtained better performance on some tests when testing students previously tested by the technician. This suggests that for some of the constituent parts of testing procedures that are more difficult for the students to understand, the nurse obtained somewhat better average performance than would be expected from a tester with similar training and experience working only with children not previously tested.

Effect of Length of School Experience. First-grade students were included in the study for the purpose of determining whether it is practicable to undertake to test children's vision before they have learned to read—which usually means before they are ready to read. It would be desirable to know whether the child has good vision before attempting to teach him to read.

For practical reasons it was not possible to test all of the first-grade students at the beginning of the school year. It is of interest to know whether the findings of the study, based on tests given at various times throughout the first year, are representative of the results that would be obtained if all testing were done at the beginning of the year. The distributions of scores obtained by students tested in different quarters of the first year have therefore been examined to see whether they show substantial differences related to the length of time the students had been in school before being tested.

Comparison of the distributions for the first, second, third, and fourth quarters of the school year shows no consistent trends. But when the distributions for the first and second half of the year are compared, it is found that on the test for far acuity by the Telebinocular, the median scores for the students tested in the second semester are fractionally higher than those for students tested during their first semester. On the tests given by the technician the median score for the first half-year was 7.8,[3] for the second half it was 8.3. On the tests given by the nurse the scores were, respectively, 6.4 and 6.9.

[3] On the Telebinocular test of acuity at far point the Snellen equivalents of scores 6, 7, and 8 are, respectively, 20/22, 20/20 and 20/18.

These differences in median scores according to semester suggest a slight improvement in performance on the test with longer school experience. If this is a real difference there is no way of distinguishing how much it is due to educational experience and how much to physiological and intellectual maturation. But the difference is so slight as compared with the differences between the median scores obtained by the 2 testers that it justifies a conclusion that the findings obtained from testing students in both semesters of the first grade are representative of what might be expected from tests made at the beginning of the school year only.

Influence of Economic Status. To determine whether there is evidence that the results obtained in the study vary in children of different economic groups, a classification of the students according to economic status has been based on the average monthly rental of homes in the city block in which the student resided. The percentile distribution of the total population of the city of St. Louis according to monthly rental paid for dwellings is used to define 5 categories of rentals. A category such as "lowest one-fifth," therefore, indicates that students in this category lived in blocks in which the average rental for homes corresponded to the average rental paid by the one-fifth of the city population paying the lowest rents.

The statistics on rentals by blocks, and the St. Louis Block Map used to find the block in which street addresses were located, were those prepared by the Metropolitan St. Louis Census Committee. The rental statistics were based on data from the 1940 census but check surveys made more recently indicate that for most areas of the city the 1940 data were still very typical of conditions at the time of the study in 1948–49.

Although in selecting schools in which to conduct the study an effort was made to choose schools that would give a cross-section of socio-economic groups in the city, the distribution according to economic status as determined by average rentals for homes in the blocks in which the students lived is quite uneven. The distribution according to 3 categories—lowest 1/5, middle 2/5, and upper 2/5—of white and Negro students is as follows:

Distribution of Students According to Economic Status

	Total [1]	Lowest 1/5	Middle 2/5	Upper 2/5
Sixth grade:				
White	457	245	98	114
Negro	132	112	18	2
First grade:				
White	527	174	137	216
Negro	67	9	57	1

[1] Information on economic status was not obtained on 17 white and 3 Negro sixth-grade students or 12 white first-grade students.

TABLE 14

Percentage of Students With Low Scores on Selected Tests [1] According to Economic Status

	Sixth Grade			First Grade		
	White		Negro	White		Negro
	Lowest 1/5	Upper 4/5	Lowest 1/5	Lowest 1/5	Upper 4/5	Middle 2/5
Number of students (100 percent)	245	212	112	174	353	57
Judgment:	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*
Clinical Judgment: Refer	27	31	40	24	21	28
M. V. T. Judgment: Refer	30	29	32	34	36	26
Far Right Acuity Tests:						
M. V. T. Less than 20/20	9	8	18	9	10	16
Tel. Less than 20/18	21	22	21	53	48	65
O.–R. Less than 20/20	22	15	18	.....	.....	.....
S.–S. Less than 20/20	27	20	21	.....	.....	.....
Heterophoria, far lateral:						
Esophoria more than / Exophoria more than						
Tel. 3.0 p. d. or 4.5 p. d.	18	24	23	14	13	24
O.–R. 3.0 p. d. or 2.0 p. d.	32	28	35	.....	.....	.....
S.–S. 2.5 p. d. or 3.5 p. d.	15	16	14	.....	.....	.....

[1] Screening tests are by technician.

The fact that in each grade most of the Negro students fall into a single category makes it impossible to compare different economic groups among them, but the study findings on the Negro groups can be compared with those for white students in the same category. Table 14 shows the percentages of students who were referred by the ophthalmologist and by the Massachusetts Vision Test, and the proportions who obtained low scores on certain parts of other procedures, according to the economic status of the students. In this table the students in the economic status groups middle 2/5 and upper 2/5 have been combined in a single group. Before this was done the distributions of scores for the 2 groups were examined separately and found to show no important differences.

In both sixth and first grades a higher proportion of Negroes was referred by the ophthalmologist than of white students in the same or other economic category. There are also somewhat higher proportions of Negro students with low scores on the tests of visual acuity than of white students. But there is no evidence of any consistent variation according to economic status in the proportions of students referred or the proportions who obtained low scores on subtests.

DISCUSSION

VISUAL FUNCTIONS cannot be measured exactly; perfect reliability is not to be expected in measurement of physiological functions. Previous studies on adults[4] of measurements of far acuity and heterophoria have shown, as might be expected, higher test-retest reliability than we obtained with children. Yet the relative reliabilities of the measures of the different functions were similar in the 2 studies. Thus, in the work with adults, most of the test-retest coefficients for visual acuity at far point and for lateral heterophoria at far and near points were between 0.80 and 0.90, while for vertical heterophoria at far and near most of the coefficients were between 0.60 and 0.70—findings that are much in line with our own.

Even the clinician cannot assume that any single measurement in his examination of an eye is exact. In the previously mentioned studies on adults, tests of visual acuity and heterophoria made with the devices used in a clinical examination were found to be no more reliable than those made with the Ortho-Rater, Sight-Screener or Telebinocular. By checking and correlating his measurements, however, the eye specialist is able to increase the reliability of the findings on which he bases his clinical judgment.

It would be useful to know the extent to which different clinicians, examining independently a sample population of school-age children, would agree on whether or not there was need for treatment. It would be of further interest to know the extent to which any disagreement resulted from real differences in their measurements, as distinct from mere differences in the standards applied.

To obtain this information would be a study in itself. In the present investigation the aim has been to provide a clinical examination that, in the view of a group of competent authorities in the field, represents best ophthalmic practice. The criteria for referral which were adopted by the team of ophthalmologists were approved in detail by the study's Ophthalmological Advisory Committee.[5] With such an eye examination,

---

[4] Project No. X–493, Bureau of Medicine and Surgery, U. S. Submarine Base, New London: Progress Report No. 2, Visual Acuity Measurements With Three Commercial Screening Devices; Progress Report No. 3, Comparative Study of Measures of Heterophoria. February 1946.

[5] The Ophthlamological Advisory Committee was composed of William L. Benedict, M. D., of Rochester, Minnesota, Chairman; the late S. Judd Beach, M. D., of Portland, Maine; Alfred Cowan, M. D., of Philadelphia, Pennsylvania; Richard C. Gamble, M. D., of Chicago, Illinois; Thomas H. Johnson, M. D., of New York City, and Lawrence T. Post, M. D., of St. Louis, Missouri.

the study has a more adequate criterion for evaluation of screening efficiency than has been available previously for any large-scale survey of vision-screening.

Adoption of other standards as a basis for clinical referrals would probably shift some students from the "refer" to the "nonrefer" category, or vice versa. But there is no reason to believe that agreement between screening results and clinical judgment would be materially affected. A lower clinical standard, for example, might reduce the number of cases "missed" by the screening procedure, but at the same time, more of the screening referrals would become over-referrals.

Good agreement between the clinical examination and the screening procedure is not to be expected as long as there is poor correspondence between their parallel parts. The index of this correspondence is the validity coefficient, which has been shown to be low in many instances (table 6). The validity coefficients are based on the full distributions of the measurements and are thus independent of any standards for referral.

Although they also reflect any unreliability of the clinical measurements, the low values of the validity coefficients clearly reflect the low test-retest correlations of the screening measurements (tables 4 and 5). As long as the screening subtests, when repeated, show poor agreement with themselves, they cannot be expected to agree with any criterion.

Since clinical judgment is not perfectly reproducible it could not be expected that a screening procedure would approach perfect agreement with the criterion. But it is reasonable to hope for a higher degree of correspondence between screening results and clinical findings than was obtained in this study. None of the procedures studied shows more than a moderate correlation with clinical judgment; no procedure refers more than half of the students found by the ophthalmologist to need eye care without giving high proportions of incorrect referrals, varying from about one-third to more than half of the total referrals.

The results of the study do not justify a conclusion that any one or two or three procedures are superior to all others for use in a screening program. As they are now set up, the Snellen and Massachusetts Vision Test give better agreement with clinical judgment than the others, but the procedures that include more measurements give a higher proportion of referrals and so miss fewer of the students who need care.

In choosing among the procedures, the administrator of a program can look at the results presented in table 2 and select the procedure with a referral rate and efficiency of measurement most nearly suited to his program. But the values shown in table 2 are likely to shift with any modifications in the administration or scoring of a procedure. For this reason the contribution made by this study lies not so much in its ranking of procedures according to relative usefulness, as in what it tells us as to why screening efficiency is not higher, and as to how efficiency can be improved either by better test construction or by better use of present procedures.

Why Is Screening Efficiency Not Higher?

It has often been pointed out that the Snellen Test misses many children who need care because it fails to identify those who have defects not involving visual acuity at far point. To overcome this limitation, other procedures have been developed that are essentially batteries of tests of different visual functions.

Great effort and ingenuity have been invested in developing these batteries of tests. Why are they not more efficient?

It has been shown in this study that the test-retest reliability of the subtests that comprise these procedures is in no case very high, and for some subtests it is very low. In other words the tests, as they stand, are subject to marked errors of measurement, at least as they are now arranged for work with children. When the referrals by several subtests having low reliability are added together, the accumulation of errors tends to defeat the purposes of adding subtests to the battery.

The question now is, why are the methods of measurement not more reliable? A likely answer would seem to be that the understandable pressure to save time in administration of the tests is the root of the trouble. There are recognized methods of increasing the reliability of measurements, but these require devotion of more time to administering the tests. The emphasis on saving time has resulted in unsuccessful attempts to short-cut the complex measurement problems involved. Hope that, even with the most ingenious instrument devisable, any visual function can be measured reliably by a quick check holds practically no promise of success, especially with children.

Is it practical, however, to consider any methods of administering the screening tests that require more time?

Referral of a higher proportion of the students who need care would in itself justify devoting some additional time to the screening tests. But from the administrative point of view the amount of time required for a screening test is an important consideration. Every effort should be made, therefore, to make the best possible use of the time involved in the screening program.

Time that can be devoted to obtaining more reliable scores on some of the subtests can be gained by eliminating from a procedure other subtests that make little or no contribution to the efficiency of the procedure as a whole. But even if the time required for administration of the screening tests must be increased in order to obtain more reliable results, there may well be compensation in the saving of professional time of nurse or teacher that would otherwise be devoted to followup of incorrect referrals.

There would also be saving of professional time if instructions for reliable methods of administering the tests can be so worked out that less highly-trained workers can be used as testers.

Would screening efficiency be higher if different standards for referral were used? The proportions of referrals could be changed, but could

better overall agreement with clinical judgment be obtained by shifting the cutoff points?

Although the general answer is "no" where only one measure or component of a procedure is concerned, the answer for a multiple-measure procedure depends on the reliability, validity, and intercorrelations of the components. Unless information about these factors is available and is carefully used in making decisions, changes in cutoffs are as likely to result in decreased efficiency as in improvement.

With the knowledge that the first part of the Massachusetts Vision Test had relatively high reliability and validity, it was possible to make a fair advance guess that lowering the cutoff of that component would decrease the procedure's overall efficiency (p. 23). In that instance consideration of the intercorrelations did not happen to be very important, but this is not always the case. The intercorrelations can affect the results of changed cutoff points in a complex manner.

Ideally, changes in the standards of a multiple-measure procedure are best decided by checking the results obtained with different cutoff points against an adequate criterion. Short of that possibility, changes are not ordinarily worth while except on the basis of expert advice, which in turn should be based upon substantial statistical information about the component parts of the procedure.

Improving Test Construction

Since greater efficiency of screening procedures depends upon finding ways of administering test materials to give high reliability and validity, and since there is reason to believe that it would be administratively feasible to devote the requisite time to such testing, those concerned with improvement of vision screening procedures should give further attention to applying certain recognized principles of test construction. Although these principles have already been employed to a considerable extent, they need to be utilized in such a way as to yield better balance in the procedures as a whole, with each of the components making a real contribution to the result.

Among the principles that can be applied to obtain greater reliability of test scores are the following:

1. If a method of measurement possesses any degree of reliability, a better approximation to true scores can be obtained by taking the average of several measurements than by using a single measurement. The number of times a measurement must be repeated or the amount of lengthening of the measuring process by other means needed to attain a specified degree of reliability can be predicted quite well, if the reliability of one administration of a test of known length is carefully determined.

   Use of the average score on tests of visual acuity may seem to be

at variance with ophthalmological practice, which is to accept the best performance obtained. But single measurements in a screening test may be presumed to be more subject to error than single measurements obtained by an expert clinician. In the screening test the aim should be to obtain a measure that is stable, or reliable. For this the average has an advantage; it is subject to less error than any single measurement.

2. Measurements made on a scale with relatively fine gradation of scores will give average scores that approximate the true scores more closely than the same number of measurements made on a scale with coarser gradations. In a screening test, precise scores are more important near the critical point on a scale of measurement than at other points on the scale. If a student's first score is far to one side of the critical point it is relatively unlikely that repetition of the test will result in a final or average score on the other side of that point. If something is already known about the distribution of scores yielded by a test in a given form, one may judge what part of the scale may need finer gradations, and also perhaps, which first scores it may be profitable to repeat.
3. A single administration of a test can often be made much more reliable by preliminary practice with unscored sample items. For some visual functions it is possible that the greater portion of the time devoted to a given test can be spent to advantage in such preliminary practice.
4. If it is found that any subtest cannot be administered in such a way as to yield a final score of satisfactory test-retest reliability without undue expenditure of time, that subtest should be omitted from the procedure. Its inclusion along with other subtests that are more reliable will usually worsen the overall efficiency of the procedure as a whole. It would also seem advantageous to omit the subtests that refer very few children who are not referred by other parts of the procedure, i. e., omit a subtest that correlates highly with one or more other subtests.
5. The value of the overall score, for referral purposes, may depend on proper weighting or interrelating of the scores obtained on individual subtests.
6. Instructions to testers should not only give precise directions as to the method of obtaining a single measurement but should include specific directions regarding practice items, repetition of tests, and methods of scoring. As little as possible should be left to the tester's judgment. Instructions should be so worked out as to assure for each subtest the maximal reliability of scores that it is practical to seek. Less specific directions, such as a general direction for the procedure as a whole to the effect that retesting is to be done under certain conditions, are not likely to be interpreted uniformly in the conditions under which school screening work often operates.

Instructions should include as an integral part of the test procedure directions for putting the child at ease in the testing situation and for helping him to understand what will be expected of him before any actual testing is started.

Instructions need to be tried out on testers typical of those who will administer the tests in school screening programs in order to determine the most useful wording and to eliminate superfluous detail. If the instructions are worked out with sufficient care, any intelligent adult who has the ability to work well with children should be able to follow them and to administer the tests successfully.

Those familiar with the various screening procedures as they are now set up will recognize that at many points considerable attention has already been given to these principles of test construction. They have not, however, been applied in a way that assures adequate reliability of each constituent part of a procedure. Without this, dependable overall scores cannot be expected.

Repetition of tests is employed in several different ways as the screening procedures are now set up.

In many school screening programs it is customary to retest all students who are referred on the first administration of the screening procedure and to refer finally only those who fail both times they are tested. As we have seen for the Massachusetts and Telebinocular Tests, this method of selecting students to be referred has some merit, but it has rather more effect in reducing total referrals than in improving screening efficiency or the correlation with clinical judgment.

Another way in which repetition is now used is to advise that when the student fails only one or two parts of the procedure, those subtests should be repeated. A second administration of these subtests may be helpful in some instances, but if the subtest is one involving many chance errors of measurement, these are almost as likely to affect the second score as the first one. And this use of repeat tests fails to take into account the fact that, through error in the first administration of the test, some students obtained initial passing scores and were therefore not referred when they should have been.

A third use of repeat tests in present testing procedures is found in instructions to the effect that a subtest is to be repeated if the tester is not satisfied that the student has given his best possible performance, or if the responses have included only a single error. Such instructions leave too much to the tester's judgment and place little emphasis on the need for any checking of passing scores. Here, too, it is apparently assumed that the results of the second test are markedly more reliable than those of the first test.

On the basis of further study of the measuring problems it should be possible to define more precisely the indications for retesting, the number of repetitions of the subtest required, and how the measurements should be combined to give the most reliable final score.

Examples of the use of sample items or practice tests are to be found now in each of the screening procedures, but their use is limited. Fine gradations of the scale of measurement are provided in a number of subtests, but the testing instructions do not provide for use of these to the greatest advantage. Much work has already gone into the development of testing instructions and this will not be lost as efforts to make the instructions more satisfactory continue.

After it has been determined which measurements are sufficiently valid and can be made reliable enough to be useful components of a screening procedure, and how the tests should be administered so as to give the required reliability, attention should be directed to the selection of appropriate standards for referral.

Ophthalmic theory and clinical experience as to the meaning of specific measurements are the basis for standards for referral, but clinical standards cannot be converted directly into screening standards. Measuring devices used for screening are usually not exactly the same as those used by the clinician, and the testers in screening programs are less skilled. Moreover, screening test cutoff points are intended to be applied arbitrarily to a series of measurements, whereas the clinician uses his judgment to evaluate one finding in the light of another. In selecting the standards for referral for a screening procedure it is therefore necessary to take into account the validity and reliability of the measurements, and the intercorrelations between component subtests.

The difficulties inherent in the application of these various factors to the selection of cutoff points are such that the process must be to a large extent one of trial and error. But it should be carried out with an awareness of the interrelationships involved and the trial should be against an adequate criterion, not, as has too often been the case, against a criterion that recognizes no errors but those of over-referral.

The many factors to be taken into account in selecting cutoff points, and the difficulties of testing adequately those selected, make the establishment of standards for referral a responsibility of those who construct the tests. The screening program administrator, and his clinical advisers, are rarely in a position to give due weight to all of the factors or to make a real evaluation of any standards they select. But since no one level of referral is likely to be suitable for all screening programs, those concerned with construction of a procedure might consider the advisability of setting up more than one set of standards for use with it, e. g., a high standard and a low standard. This would reduce the likelihood that program administrators would be tempted to introduce untested modifications.

More Effective Use of Present Tests

The discussion thus far has related to how screening procedures might be rebuilt to make them more efficient. What answers does the study

give to the school health administrator who wants to plan a program with existing materials?

Whatever testing device is employed, its efficiency will depend on its being used in a way that will give reliable test scores. There is no equipment that will give good screening results under poor testing conditions or with hurried, unchecked testing.

First of all, school administrators should share with the tester responsibility for seeing that the best possible screening conditions are provided, in accordance with the instructions for the particular procedure to be used.[6] Administrators and testers should also work together in scheduling the testing so it can be conducted in unhurried fashion.

Second, a plan of repeating the tests should be adopted. Pending the time when methods of administering the tests to give dependably reliable results have been worked out, the following plan is suggested:

Administer the screening procedure to each student twice. Preferably the second test will not immediately follow the first, but it may be given after an interval of at least 10 minutes, or perhaps the next day. Whenever the scores thus obtained *on any subtest* are contradictory—one "pass" and one "fail"—that subtest should be repeated. The score obtained two out of three times will be the final score for the subtest. Suitably averaging the first and second scores may be substituted for the two-out-of-three plan if the subtest is one that has a finely graded scale.

One or two repetitions of each subtest should be the minimal requirement. Higher reliability can be achieved by averaging a larger number of scores. Repetition should be used to increase the reliability of each individual subtest rather than as a check on the overall refer-nonrefer result of the procedure.[7]

Attention to these aspects of testing is essential whatever screening procedure is employed.

The study has shown that the Snellen Test gives as good agreement with clinical judgment as any of the multiple-test procedures, and better than most of them, though its referral rate is lower. Many school health programs do not have facilities for adequate followup of more students than the Snellen will refer. In this situation little is to be gained, and there may well be some loss, in adopting a procedure that gives a larger number of referrals, more of which will be over-referrals. If a similar consideration seems to lead to choice of the low standard for the Snellen rather than the high standard, such a choice should take into account not only the lower over-referral rate of the low standard but its much lower rate of correct referrals. It refers only about a fourth of the students who need care, whereas the high standard refers about half.

What has been said of the Snellen Test would apply equally to inde-

[6] For instructions for testing conditions for the Snellen Test see: A Guide for Eye Inspection and Testing Visual Acuity, Publication 180, National Society for the Prevention of Blindness, 1790 Broadway, New York 19, N. Y. Price $.05.

[7] For footnote 7 see page 55.

pendent use of the tests of far acuity included in the Massachusetts and Telebinocular Tests and, at least for upper grade students, those in the Ortho-Rater and Sight-Screener.

School health programs, with better-developed facilities for followup, may not be content with a screening program that does not refer more than half of the students who need care, so will prefer one of the multiple-test proccdures. As they are now set up, the Massachusetts Vision Test is the most efficient of these. But the efficiency of any of them can be improved by proper use of retesting, and by dropping some of the subtests that are shown to have low reliability or validity—for example, tests of fusion and tests of vertical heterophoria, at far or near point. Time gained by omission of these subtests could be used for repetition of the remaining parts of the procedure. Such a modification of one of the multiple-test procedures can be adopted for use until those concerned with improvement of screening procedures provide better-tested and more precise instructions. (Modification of the procedures by shifting the cutoff points is not recommended unless it can be based on consideration of the relative validity and reliability of the subtests and the intercorrelations among them, as has been discussed elsewhere.)

For first-grade students, the low reliability and validity for nearly all subtests except that for far acuity makes it appear unlikely that anything is to be gained by using one of the existing multiple-test procedures with young children. For present use the Snellen alone, or possibly the Massachusetts Vision Test, if higher referrals are acceptable, is probably to be preferred below third or fourth grade.

[7] The following example illustrates use of repeat tests as a check on the overall score as compared with their use to increase the reliability of individual subtests:

	First administration of test	Second administration of test	Third administration of test	Average score
Subtest A	6	7	6	6
Subtest B	7	6	7	7
Subtest C	5	6	7	6
Subtest D	8	8	8	8
Subtest E	6	6	5	6
Overall scores	Refer	Nonrefer	Refer	Nonrefer

If it is assumed that a score of 5 or less means failure of subtest, and failure of 1 or more subtests is cause for referral, then the student has an overall score of "refer" 2 out of 3 times.

But if the repeat tests are used to increase the reliability of scores for individual subtests, then the scores obtained on each subtest will be averaged. For no subtest is the average score as low as 5, so the final overall score is "nonrefer." It is clear that "nonrefer" is the sounder of the 2 possible interpretations of this student's performances.

Educators have developed reliable methods of tests and measurements in their own field. If they recognize that the same principles apply to measurement of a physiological function, they will appreciate the need for having proper testing conditions and for allowing time for accurate measurement in order to obtain dependable results from a vision screening procedure. Time devoted to obtaining reliable scores means referral of more of the students who need care and saving of time spent needlessly on followup of over-referrals.

Since there are some abnormalities of the eye that will be missed by any testing device, teacher observation, checked, when possible by school nurse or physician, should be used as a supplement to any vision-testing program. In the presence of persistent signs or complaints suggestive of eye trouble, it should not be assumed that ability to pass a screening test eliminates the need for a thorough, professional eye examination.

APPENDIX A

History and Personnel of the Study

FOR MANY YEARS the National Society for the Prevention of Blindness has been concerned with the problem of selection of screening procedures that would best identify those children who have visual defects.

In 1929 the Society conducted a study that resulted in the recommendation of a battery of tests for use at the preschool level. In 1939 the Society appointed a Committee on Vision Testing Procedures. This group reviewed and summarized the most pertinent studies on visual screening of children that had been published between 1924 and 1939. The group also recommended that a study be made to determine suitable testing procedures including indoctrination of examiner personnel. Although the advisability of such a study was recognized, no funds were available at that time.

In 1943 the National Society called a conference of administrators of the various agencies concerned with the prevention of blindness and related health services. The topics for discussion were problems relating to the visual screening of preschool and school children. The consensus was that testing and followup were inadequate and not well standardized. It appeared obvious, too, that test procedures had not been properly validated.

The result of the meeting was a reorganization in 1944 of the Committee on Visual Testing Procedures as the Advisory Committee on Visual Screening Programs. An attempt was made to secure for this committee representation from all interested groups. A pediatrician from the Children's Bureau (of the Federal Security Agency, which has since become the Department of Health, Education, and Welfare) who had served previously on the original committee was made chairman. This new committee submitted to the National Society a recommendation that the Society sponsor research on methods of visual screening. Similar recommendations were made to the National Society in 1946 by the secretary of the School Health Section of the American Public Health Association.

Through its representative on the National Society's committees the Children's Bureau had participated in the formulation of the recommendations made by the committees. In 1943 the Bureau made preliminary plans for a study of vision testing procedures but the national situation was such that no funds could be made available at that time.

After the pressures of World War II had subsided the Children's Bureau and the National Society for the Prevention of Blindness developed a tentative plan for a study to be conducted as a joint project of several interested agencies. This proposal was presented for the consideration of a conference called by the National Society in June 1947. Attending this conference were representatives of the fields of health, education, and welfare. A nucleus of the group were members of the National Conference for Cooperation in Health Education. The group agreed unanimously on the need for a study of vision testing procedures for use in elementary schools and made many helpful suggestions regarding the planning of such a study.

The tentative plan for the study called for various types of participation by State and local agencies, and it was found that in St. Louis this cooperation would be available to a generous degree.

The Division of Health of the Missouri State Department of Public Health and Welfare was prepared to join in the sponsorship of the project. The St. Louis Board of Education approved the participation of the public elementary schools of the city in the study. It was necessary that the project be conducted in a large school system with a well-organized school health program such as exists in St. Louis. The Department of Ophthalmology of the Washington University School of Medicine undertook administration of the clinical eye examinations that would be an essential part of the study. The Department's Director of Graduate Training in Ophthalmology agreed to serve as Ophthalmological Director of the project, and the Office of Naval Research approved his inclusion of the study as a part of the research program for which they were giving him support. The various types of cooperation thus available were determining factors in the selection of St. Louis as the site for the project.

The plan also called for a project that would be national in scope in the sense that it would be designed and conducted in consultation with a group of ophthalmologists from different parts of the country who have achieved national recognition as authorities in the field. An Ophthalmological Advisory Committee of six such authorities was therefore appointed. This Committee approved the design of the study, with special attention to the content of the ophthalmological examination and interpretation of the clinical findings, and kept in touch with the progress of the program.

An Executive Committee composed of representatives of the principal sponsors of the project was responsible for administration of the project. A complete list of personnel participating in the study is shown below.

The program in the schools was begun in February 1948, and completed in May 1949.

All associated with the study regretted the illness and untimely death in June 1952 of its Ophthalmological Director, Dr. Richard G. Scobee. He contributed to the study the benefit of his background of scientific research in ophthalmology as well as his clinical knowledge of the field.

Personnel of the Study

The Executive Committee is composed of:

1. Marian M. Crane, M. D.
   Chief, Research Interpretation Branch, Children's Bureau.
2. Franklin M. Foote, M. D.
   Executive Director, National Society for the Prevention of Blindness.
3. L. Marion Garner, M. D.
   Director, Division of Child Hygiene, Missouri Department of Public Health and Welfare.

The Ophthalmological Advisory Committee is composed of:

Chairman: William L. Benedict, M. D., Professor of Ophthalmology, Mayo Foundation Graduate School of Medicine: Executive Secretary, American Academy of Ophthalmology and Otolaryngology.

Lawrence T. Post, M. D., Professor of Ophthalmology, Washington University School of Medicine.

Richard C. Gamble, M. D., Senior Attending Ophthalmologist, St. Luke's Hospital; Attending Ophthalmologist, Children's Memorial Hospital, Chicago.

Thomas H. Johnson, M. D., Associate Clinical Professor of Ophthalmology, College of Physicians and Surgeons, Columbia University.

Alfred Cowan, M. D., Professor of Ophthalmology, Post Graduate School of Medicine, University of Pennsylvania.

Sylvester Judd Beach, M. D., Secretary-Treasurer, American Board of Ophthalmology.

School Authorities:

The cooperation of the public school system in St. Louis left nothing to be desired. Particular thanks are due the following:

Mr. Philip J. Hickey, Superintendent of Instruction.

Mr. Edward H. Beumer, Assistant Superintendent in Charge of Elementary and Special Schools.

Mr. Clement A. Powers, Assistant Director of Education, Division of Tests and Measurements.

Lloyd L. Tate, M. D., Director, Division of Health and Hygiene.

Miss Mary E. Stephenson, Supervisor of Nurses, Division of Health and Hygiene.

Directing Staff:

Ophthalmological Director: Richard G. Scobee, M. D., Assistant Professor of Ophthalmology, Washington University School of Medicine.

Directing Staff—Continued

Statistical Director: Earl L. Green, Ph. D., Associate Professor of Zoology, Ohio State University, Columbus, Ohio; now Geneticist, Biology Branch, Division of Biology and Medicine, United States Atomic Energy Commission.

Study Administrator: Marian M. Crane, M. D., Chief, Research Interpretation Branch, Children's Bureau.

Assistant Study Administrator: Marguerite Furey, R. N., Consultant in Nursing Activities, National Society for Prevention of Blindness (until April 1948).

Helen E. Weaver, R. N., Consultant in Nursing Activities, National Society for Prevention of Blindness (from April 1948).

Deputy Study Administrator: Ann de Huff Peters, M. D., Medical Research Assistant, Division of Research in Child Development, Children's Bureau.

Ophthalmologists (Washington University School of Medicine):

Richard G. Scobee. M. D., Assistant Professor of Ophthalmology.

David M. Freeman, M. D., Assistant Chief Resident in Ophthalmology.

Arthur W. Stickle, Jr., M. D., Fellow in Ophthalmology.

George T. Stine, M. D., Chief Resident in Ophthalmology.

Nurse Coordinator: Annette L. Gronemeyer, R. N., Division of Health and Hygiene, St. Louis Public Schools.

Technician: Mr. Raymond G. Stratmann, National Society for Prevention of Blindness.

Administrative Assistants (Missouri State Division of Health):

Mrs. Claude E. Stephens (until May 1948).

Mrs. Caroline H. Holzum.

Mrs. Lesta Ferguson.

Miss Mildred A. Schroeder.

Mrs. Mardel Tivener (January to March 1949).

Special Consultants:

Statistics:

Bronson Price, Ph. D., Public Welfare Research Analyst, Technical Studies Branch, Division of Research, Children's Bureau.

Eleanor P. Hunt, Ph. D., Acting Chief, Program Analysis Branch, Division of Research, Children's Bureau.

Edward B. Olds, Ph. D., Research Director, Social Planning Council of St. Louis and St. Louis County.

C. Edith Kerby, Statistician, National Society for the Prevention of Blindness.

Medical Social Service:

Miss Ruth C. Olson, Regional Medical Social Work Consultant, Division of Health Services, Children's Bureau.

Participating Schools:

Fanning, Mr. C. E. Stephens, Principal.
Mullanphy, Mr. W. H. Schleuter, Principal.
Clark, Mr. W. J. See, Principal.
Lincoln, Mr. Lucian P. Garrett, Principal.
Clinton-Peabody, Mr. Stephen L. Pitcher, Principal.
Humboldt, Dr. Wm. Hall Todd, Principal.
Webster, Mr. Logan R. Fuller, Principal.
Baden, Mr. Howard E. Green, Principal.
Laclede, Mr. Fred S. Milan, Principal.
Hamilton, Miss Percy A. Lyon, Principal.
Dewey, Mr. Leo P. Granger, Principal.
Hempstead, Miss Susan B. Ryan, Principal.
Emerson, Miss Ethel Wurdack, Principal.
Cote Brilliante, Mr. John M. Langston, Principal.

School Nurses:

Fanning, Mrs. Viola Farmer.
Mullanphy, Miss Kay O'Donnell.
Clark and Hamilton, Mrs. Thelma McCann.
Lincoln, Miss Ethel Howard.
Clinton-Peabody, Miss Maxine Brandt.
Humboldt, Miss Emily Schott.
Webster, Miss Agnes Cosgrove.
Baden, Miss Vera Miessner.
Laclede and Hempstead, Miss Frances Griffith.
Dewey, Mrs. Pearl Gilsdorf.
Emerson, Miss Mary Goldie Watta.
Cote Brilliante, Miss Pauline Craig.

School Teachers:

Fanning:
Miss Julia Schmidt.
Miss Eugenia Henke.
Miss Dorothea M. Galvin.

Mullanphy:
Miss Lula Hack.
Miss Marie Lyons.
Miss Margaret McCormick.
Mrs. Irene Kelly.

Clark:
Miss Mollie Cotler.
Miss Gertrude R. Davis.
Miss Martha Leonard.

Lincoln:
Miss Nannie E. Jones.
Miss Elizabeth Givens.
Miss Clara Washington.
Miss Ida Jones.
Miss Alma Loving.
Mr. Ogie Wilkerson.

Clinton-Peabody:
Miss Agnes Mohan.

Miss Alice Nerlich.
Miss E. C. Baker.
Miss Shirley Lyon.
Miss Kathryn Frei.
Miss Esther Dornhoefer.
Miss Mildred Erskine.

Humboldt:
Miss Anna Marie Lottmann.
Miss Lucille Huger.
Mr. Richard M. Jentsch.
Miss Helen Anderlan.
Miss Elizabeth Schwarz.
Miss Edna Murphy.

Webster:
Miss Elizabeth M. Bick.
Mrs. Ruth Golden.
Mrs. Irene Mayer.
Miss Mae Schulte.
Miss Dorothy Horan.
Miss Marjorie Murrin.
Miss Celine Lawrence.
Miss Laura d'Arcambal.

Baden:
Miss Carlene Keller.
Miss Winifred Hosch.
Miss Helen Kelly.
Miss Valentina Marco.
Miss Margaret Weaver.

Laclede:
Miss Cary H. Randolph.
Miss Sallie Leonard.
Miss Joan McMullen.
Miss Evelyn Schultz.

Hamilton:
Miss Marguerite B. Hallam.
Miss Hilda Hageman.
Miss Cathryn Liebig.

Dewey:
Miss Lucille Boulicault.
Miss Mary B. Womack.

Hempstead:
Miss Elsie N. Dodson.
Miss Ruth Cornelius.
Miss Dorothy Zimmerman.

Emerson:
Miss Agnes Staed.

Cote Brilliante:
Miss Marguerite Stewart.
Miss Grace L. James.
Miss Janet Thompson.

All of those cooperating in the study have been listed. With their help, the framework was laid and actual testing begun in February 1948; it was completed in May 1949.

APPENDIX B

Standards for Referral

Standards for Referral by Clinical Examination: An eye specialist's decision as to whether or not an individual needs treatment for his eyes depends upon his judgment of the significance of his measurements of visual function in relation to each other and in relation to the presence or absence of symptoms or evidences of eye pathology. There are no hard and fast limits which automatically classify as in need of treatment every individual who has measurements outside these limits. The clinician does have in mind, however, a range of measurements usually associated with good visual function and gives special attention to the significance of any measurement falling outside this range. For convenience in discussion, this range of measurements will be termed the "normal range," and its limits called the "limits of normal," but this terminology is not to be interpreted as implying that there are actually any precise limits to the range of normal measurements.

For the purposes of this study it was necessary to define limits of normal for the measurements included in the clinical examination. The definitions were needed as a means of assuring a uniform basis for clinical judgments and as a means of communicating to others the basis on which the clinical judgments were formulated.

There is lack of unanimity among eye specialists as to the range of normal for the various visual functions at different ages. The limits established for this study are those approved by a group of ophthalmologists qualified by clinical knowledge and experience to make decisions of policy in this area. The limits were defined tentatively by the Ophthalmological Director of the study and adopted after being approved by the Ophthalmological Advisory Committee.

The range of normal for clinical measurements as thus defined is as follows:

Inclusive Limits of "Normal" for Clinical Measurements

Sixth Grade

Visual acuity	20/20 or better in each eye and equal acuity in both eyes.
Lateral heterophoria, far:	
Maddox Rod	6 p. d. exophoria—8 p. d. esophoria.
Cover Test	5 p. d. exophoria—5 p. d. esophoria.

Lateral heterophoria, near:	
Maddox Rod	5 p. d. exophoria—6 p. d. esophoria.
Cover Test	8 p. d. exophoria—6 p. d. esophoria.
Maddox Wing	6 p. d. exophoria—5 p. d. esophoria.
Vertical heterophoria, far:	
Maddox Rod or Cover Test	1 p. d. right—1 p. d. left hyperphoria.
Vertical heterophoria, near:	
Maddox Rod, Cover Test, or Maddox Wing	1 p. d. right—1 p. d. left hyperphoria.
Spherical equivalent	zero through +3.00 diopters

First Grade

The following limits of normal for first grade are the only ones that differ from those shown for sixth grade:

Visual acuity	20/30 or better in each eye and equal acuity in both eyes.
Spherical equivalent	zero through +3.50 diopters

These limits of normal were used as guides by the ophthalmologist in forming his clinical judgment, but a measurement falling outside these limits did not arbitrarily determine that the judgment would be "refer." The significance of such a measurement was evaluated in relation to other findings.

Pathology or congenital anomaly of the fundus, or external pathology, was considered cause for referral if treatment was indicated.

In a preliminary evaluation of the clinical findings students were classified in 3 groups: "true referrals," "theoretical referrals," and "nonreferrals." The "theoretical referrals" were those students who had some finding that could not be properly evaluated without a complete ophthalmological examination but who were found, when the examination was completed, not to be in need of treatment for their eyes. Most of these were students who had complaints, signs of external pathology, or subnormal visual acuity measurements when first tested, but no significant refractive errors. A few were students who had actual visual defects but whose defects could not be corrected by treatment. This group of "theoretical referrals" was combined with the "nonreferrals" for comparison of the results of the screening procedures with the clinical judgment, since the usefulness of a screening procedure depends upon its success in selecting those students who actually need treatment.

Standards for Referral by Screening Procedures: A screening test gives a series of measurements or scores. To use the test for selection of individuals who are to be referred for eye care it is necessary to apply standards for referral, or cutoff points that define measurements that are to be regarded as satisfactory and measurements that are to be considered indicative of need for referral.

For the Snellen Test, 2 standards that are frequently employed in school screening programs have been used in this study. These are designated as the "high standard" and the "low standard." According to the "high

standard" a student is referred if with either eye he fails to read the 20/20 line correctly; according to the "low standard" he is referred if with either eye he fails to read the 20/30 line or better. Standards based on the score obtained for "the better eye" were not used because unequal acuity in the two eyes may in itself be an indication of need for an eye examination.

Two standards for referral were likewise applied to the scores obtained with the Near Vision Test. The "high standard" refers any student who with either eye fails to read the 14/14 line: the "low standard" refers him only if he fails to read the 14/17 line or better.

With combinations of screening procedures, such as the combination of the Snellen and Near Vision Tests, a student is classed as a referral if he is referred by any test included in the combination.

The Massachusetts State Department of Public Health has defined the standard for referral for the Massachusetts Vision Test and this standard is used in the study.

For the Telebinocular Test the standard for referral recommended by the manufacturer in the 1947 revision of the Manual of Instructions [8] is used. A student is referred if he has one or more measurements outside the range shown as "expected" or "doubtful" on the record forms supplied by the manufacturer. In accordance with the recommendation of the manufacturer, the results of "Test 1: Simultaneous Vision" are ignored in determining need for referral.

A second interpretation of the Telebinocular measurements is based on a standard developed in the same way as the standards for the Ortho-Rater and Sight-Screener. This has been designated the "study" standard to distinguish it from the "manufacturer's" standard.

It seemed desirable in this study to find out how efficient the total Telebinocular Test is for first-grade students, but it should be noted that this is not the use of the

---

[8] The Manual of Instructions for the Telebinocular (1947 revision) in use at the time of the study states: "One check mark in the undesirable column indicates that the pupil should be referred to an eye specialist." This instruction has been changed in the 1952 revision to: "One check mark in the undesirable area indicates that the tests should be remade immediately to check error in response or in question by the operator. Continued failure on this test should be watched and re-checked every 6 months. Two or more check marks in the undesirable columns indicate need for an immediate referral. Likewise one check mark in an undesirable column if it refers to usable vision—either far- or near-point—indicates need for referral. Failure to see 3 balls or seeing 3 balls quickly changing to 4 in Tests 4 and 11 (Fusion) always indicates referral."

In this study all test scores have been interpreted on the basis of the instructions in effect at the time of the study. The less specific instructions for referral in the 1952 revision of the manual could not have been carried out fully in a study of this type even if they had been in effect at the time of the study. Difficulties in the application of test instructions of this kind are considered in the Discussion (pp. 51–52).

The application of arbitrary cutoff points to all scores obtained on subtests of each screening procedure assures objectivity in evaluation of the different procedures.

procedure that the manufacturer recommends for this age group. The Manual of Instructions says: ". . . young children who fail the Visual-Survey Tests of lateral imbalance, fusion, or near-point usable vision may be visually immature rather than visually deficient. In keeping with the recommendations of many educators, the school should require only far-point reading activities from all 6- and 7-year-old children who fail one or more of the near-point tests. Children who fail the acuity tests at both far and near points should be referred to an eye specialist for attention."

At the time of the study the manufacturers of the Ortho-Rater and the Sight-Screener had made no recommendations as to how the measurements obtained with these instruments should be interpreted. It was, therefore, necessary to establish the standards for referral by these tests that would be used in the study.

One method of setting standards for referral is to apply directly to the screening test the limits of normal as defined clinically. This does not take into account the possibility of differences in calibration of the scales of measurement of the clinical test and the screening test of the same function. Nor does it allow for any consistent tendency for screening testers to obtain higher or lower readings than the clinician, even though using the same test. Such a difference might result from such factors as greater skill on the part of the more experienced tester in eliciting maximal performance, or better performance of students to the screening tester who is a familiar person in a familiar environment.

These difficulties are reduced if the standards for referral are based on the actual distributions of clinical and screening test measurements made on a typical sample of the population concerned.

Application of this principle is best described by an example: Let it be assumed that for fifth-grade students the lower limit of normal visual acuity has been defined clinically as 20/20 and that clinical examination of a random sample of fifth-grade students finds that 30 percent have scores below this limit. And let it be assumed that on the same or a similar sample of students a screening test of visual acuity finds only 15 percent with scores below 20/20, but finds 30 percent with scores below 20/18. The standard for referral for this screening test would be set to require referral of all students with scores of less than 20/18 on the screening test scale.

This method of setting standards depends upon knowing the distribution of measurements by each test in the populations concerned. Such data have not been available previously, but in this study measurements were obtained on large groups of sixth- and first-grade St. Louis public school children who were unselected insofar as possible. The distributions of measurements obtained on these groups in the clinical examination and in the screening tests as administered by the technician have been used for the Ortho-Rater and Sight Screener and the Study Standard for the Telebinocular. It is recognized that these distributions were imperfect for the purpose, but they are better than have previously been available.

The standards were established quite independently of *overall*

clinical judgment as to whether or not students needed eye care, and according to a procedure adopted before any actual data became available.

Below are further details on the methods used to set standards for referral on certain measures.

Far visual acuity was tested twice in the clinical examination if the first test indicated subnormal acuity. The second test, after the student was more accustomed to the examination procedure, found considerably fewer individuals with subnormal visual acuity: 19 percent of sixth-grade students and 9 percent of first-grade students. Standards of referral for screening tests of visual acuity at far point are based on these distributions. In the absence of a test for visual acuity at near point in the clinical examination, the same distributions were used as a basis for setting standards for referral for screening tests of near visual acuity.

The distributions for far and near vertical heterophoria are unusual because, both on clinical examination and in the screening tests, very few measurements depart from modal performance on the test as it is given and scored. The clinically defined limits of "normal" are therefore used as the limits of measurements considered satisfactory for nonreferral by the screening tests.

No measurement of stereopsis or depth perception was included in the clinical examination so the standards for referral for screening tests of this function are based entirely on the distribution of scores obtained on those tests. Selection of appropriate cutoff points was based on 2 considerations: (1) If the measurements showed a "natural break" at a reasonable point in the distribution (i. e., a piling up of individuals with measurements on one side of that point and relatively few individuals on the other side) that point suggested a limit for the usual—and presumably "normal"—measurements. (2) In the absence of other evidence, the most extreme 5 percent of measurements might be taken to represent unusual and presumably "abnormal" measurements.

For the tests of fusion and binocular vision included in some of the screening procedures the method of scoring does not represent a graduated scale of measurements. The standards for referral selected are those that will refer individuals who, according to the test, do not have simultaneous binocular vision or who cannot achieve fusion.

Each screening procedure is composed of a series of measurements. As the standards for referral have been applied in the study, a student is referred if one or more of the measurements is outside the limits considered acceptable for nonreferral. This means that a subtest whose standard is set relatively high, i. e., to refer a large proportion of students, receives more weight in deciding referral or nonreferral than a subtest whose standard is set to refer only a few students. Since this is the only system of weighting customarily used in screening procedures, it is the only system employed in this study.

The standards for referral by screening procedures used in the study are shown below in terms of scores accepted as evidence of satisfactory performance and, as necessary, in terms of their equivalents in comparable absolute values (Snellen equivalents or prism diopters) or other description of performance achieved. Absolute equivalents for measurements of stereopsis are available only for the Sight-Screener test of stereopsis, which is based on the Shepard-Fry Scale.

If a student obtains on any part of a procedure a score outside of the range shown as satisfactory, he is classed as a referral by that procedure.

MEASUREMENTS ACCEPTABLE FOR NONREFERRAL

SNELLEN TEST	High Standard	Low Standard
Right eye	20/20	20/30 or better
Left eye	20/20	20/30 or better

NEAR VISION TEST	High Standard	Low Standard
Right eye	14/14	14/17 or better
Left eye	14/14	14/17 or better
Both eyes	14/14	14/17 or better

MASSACHUSETTS VISION TEST

E symbols	
Right eye	20/20
Left eye	20/20
Plus sphere	Unable to read 20/30 line with either eye; or, unable to read 20/30 line with one eye, reads 20/30 line but unable to read 20/20 line with other eye.
Muscle balance	
Far vertical	Through window
Far lateral	Through house
Near lateral	Through panel

ORTHO-RATER	Measurement	Absolute Value
Far Point		
1 Phoria, vertical	3–8	1.0 left hyperphoria—1.0 right hyperphoria [1]
2 Phoria, lateral	2–13	6.66 esophoria—4.33 exophoria [1]
3 Acuity, both	10–15	20/20 or better
4 Acuity, right	10–15	20/20 or better
5 Acuity, left	10–15	20/20 or better
6 Depth	B–H	[2]
Near Point		
1 Acuity, both	10–15	20/20 or better
2 Acuity, right	10–15	20/20 or better
3 Acuity, left	10–15	20/20 or better
4 Phoria, vertical	3–8	1.0 left hyperphoria—1.0 right hyperphoria [1]
5 Phoria, lateral	3–14	7.5 esophoria—9.0 exophoria [1]

SIGHT-SCREENER		
Far Point—Red Series		
1 Binocular vision	4 letters	Simultaneous binocular vision
2 Acuity, right	20/20–20/10	20/20 or better
3 Acuity, left	20/20–20/10	20/20 or better
4 Acuity, both	20/20–20/10	20/20 or better
5 Depth	D–E	90%–105%
6 Muscle balance, vertical	2–6	1.0 right hyperphoria—1.0 left hyperphoria [1]
7 Muscle balance, lateral	8–20	7.0 esophoria—5.0 exophoria [1]
Near Point—Black Series		
1 Binocular vision	4 letters	Simultaneous binocular vision
2 Acuity, right	20/20–20/10	20/20 or better

3 Acuity, left	20/20–20/10	20/20 or better
4 Acuity, both	20/20–20/10	20/20 or better
5 Depth	D–E	90%–105%
6 Muscle balance, vertical	2–6	1.0 right hyperphoria—1.0 left hyperphoria [1]
7 Muscle balance, lateral	7–24	8.0 esophoria—9.0 exophoria [1]

TELEBINOCULAR — **Manufacturer's Standard**

Far Point		
2 Vertical imbalance	1–1	0.5 right hyperphoria—0.5 left hyperphoria [1]
3 Lateral imbalance	8–10	1.75 esophoria–3.25 exophoria [1]
4 Fusion	3, or 4 then 3	Fusion achieved
5 Usable vision right	5–10	20/25 or better
6 Usable vision left	5–10	20/25 or better
7 Stereopsis	9–12	(2)
Near Point		
10 Lateral imbalance	4–6	2.81 esophoria—4.70 exophoria [1]
11 Fusion	3, or 4 then 3	Fusion achieved
12 Usable vision right	13–22	20/25 or better
13 Usable vision left	13–22	20/25 or better
14 Usable vision both	13–22	20/25 or better

Study Standard—Sixth Grade

Far point		
2 Vertical imbalance	2–2	1.0 right hyperphoria—1.0 left hyperphoria [1]
3 Lateral imbalance	7–11	4.25 esophoria—5.75 exophoria [1]
4 Fusion	3, or 4 then 3	Fusion achieved
5 Usable vision right	8–10	20/18 or better
6 Usable vision left	8–10	20/18 or better
7 Stereopsis	10–12	(2)
Near Point		
10 Lateral imbalance	3–7	6.56 esophoria—8.44 exophoria [1]
11 Fusion	3, or 4 then 3	Fusion achieved
12 Usable vision right	14–22	20/22 or better
13 Usable vision left	14–22	20/22 or better
14 Usable vision both	14–22	20/22 or better

Study Standard—First Grade

The same as sixth grade study standard with the following exceptions:

Far Point		
5 Usable vision right	4–10	20/28 or better
6 Usable vision left	4–10	20/28 or better
Near Point		
12 Usable vision right	11–22	20/28 or better
13 Usable vision left	11–22	20/28 or better
14 Usable vision both	11–22	20/28 or better

[1] Muscle balance (phoria) is measured in prism diopters.

[2] Absolute equivalent not available.

APPENDIX C

Distributions of Measurements

TABLES 15–36 show the distributions of scores obtained in the measurements that comprise the clinical examination and the screening tests. The screening test scores are those obtained by the technician.

Each table shows, for a given visual function, the distributions of all measurements made of that function in the clinical examination or screening procedures.

The distributions of scores by each method of measurement are shown separately for students who were classed as referrals by the ophthalmologist and for those whom he classed as nonreferrals. To save space the total for each pair of distributions is not shown. However, where the relationship of the scores to the criterion (ophthalmologist's referral or nonreferral) is not of special interest, any of the pairs of distributions may easily be added together to obtain the distribution of all scores.

Crosslines through the columns represent the pass-fail cutoff points according to the standards for referral used in the study. This makes it possible to see how many measurements are "passing" scores and how many are "failing" scores among the students referred by the ophthalmologist and among those whom he did not consider in need of referral. Since more than one standard for referral was used on some tests, solid or broken crosslines are used to indicate cutoff points according to the different standards, as follows: Study standard _____; Manufacturer's standard . . . . . ; High standard – – – – –; Low standard . — . —.

Since score values on the scales of measurement often differ from one procedure to another, it was necessary to do a certain amount of rounding of the score values at some points in the distributions. For example, Telebinocular lateral heterophoria score equivalents include fractional values, but in the tabulations they are rounded to the nearest whole prism diopter.

The tests, as given and scored, often do not measure performance beyond a certain point on the scale, but use an "open-end" interval which, in effect, lumps all performance at or beyond that point into one category. Therefore, in some instances the tabulations make it appear that a score at the extreme end of a scale represents a particular value, whereas it actually represents the value shown and all measurements beyond. For example, in table 21, showing the distribution of far lateral heterophoria measurements on sixth-grade students, the table seems to show that on

TABLE 15

FAR ACUITY, MONOCULAR

Sixth Grade

Cutoff points for the various standards for referral are shown as follows: Study std. ——; Manufacturer's std.: .....; High std.: – – –; Low std.: –.–.–.–.

	Clinical Examination		Ortho-Rater		Sight-Screener		Telebinocular		Mass. (Snellen)	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	190	419	190	419	190	419	190	419	190	419
Right eye.. Poorer than 20/20 ...	118	69	74	39	88	54	70	29	60	7
Right eye.. 20/20 or better.....	72	350	116	380	102	365	120	390	130	412
Left eye... Poorer than 20/20 ...	126	71	87	52	79	66	74	25	73	11
Left eye... 20/20 or better.....	64	348	103	367	111	353	116	394	117	408
Right and Left Eyes Combined										
Total measurements:	380	838	380	838	380	838	380	838	380	838
20/100 or poorer	57	2	11	2	14	1	33	8		
20/50 to 20/70	50	4	8	1	22	4	36	3		
20/40	34	5	26	1	19	4	31	5	[2] 63	[2] 5
20/35	([1])	([1])	19	2	([1])	([1])	([1])	([1])	–.–.–.–.–	
20/30	27	17	37	2	112	111	([1])	([1])	70	13
20/28	([1])	([1])	([1])	([1])	([1])	([1])	18	10		
20/25	76	112	33	21	([1])	([1])	15	11		
20/22	([1])	([1])	27	62	([1])	([1])	11	17		
20/20	97	395	38	76	137	329	30	52	247	820
20/19	([1])	([1])	70	255	([1])	([1])	([1])	([1])		
20/18	([1])	([1])	([1])	([1])	([1])	([1])	49	109		
20/17	([1])	([1])	82	289	([1])	([1])	109	399		
20/15 or better	39	303	29	127	76	389	48	224		

[1] Score value not on the scale as it was used.

[2] Poorer than 20/30.

For the clinical examination, distributions shown in the table are for measurements of acuity without glasses. In the screening tests, however, students who had glasses were tested with glasses. The distribution of clinical acuity measurements made with glasses, if the student had glasses, is the same for students not referred by the ophthalmologist as that shown in the table; for students referred by the ophthalmologist it is as follows: 20/100 or poorer, 40; 20/50 or 20/70, 48; 20/40, 34; 20/30, 31; 20/25, 78; 20/20, 108; 20/15 or better, 41.

the Ortho-Rater test 19 students had 8 p. d. of esophoria, whereas the score actually represents 8 p. d. or more.

With a few exceptions, the distributions shown for sixth grade are for measurements on 474 white students and 135 Negro students. Those for first grade are for 539 white students. In the preliminary analysis of the study data, all data for white and Negro students were tabulated separately and the number of Negro first-grade students studied was so few it did not justify detailed analysis. Consequently, when it was decided later to combine the findings on white and Negro students, some of the detailed breakdowns had not been tabulated for the first-grade Negro children. Since that group constitutes only 11 percent of the first-grade students examined, its inclusion or exclusion has little effect on the distributions of the measurements obtained.

The instances in which distributions are for fewer than 609 sixth-grade or 539 first-grade students are certain clinical tests. Measurements of near point convergence and of prism convergence and divergence were made routinely in the clinical examination until enough data had been obtained to show the typical distribution of such measurements. After that vergence power was determined only occasionally when necessary to evaluate the significance of an unusual degree of heterophoria. Measurements of muscle imbalance at near point with the Maddox wing were also discontinued on first-grade students after enough data had been ob-

TABLE 16

FAR ACUITY, BINOCULAR

Sixth Grade

Cutoff point for the Study standard for referral is shown by line across columns.

	Ortho-Rater		Sight-Screener	
	Ref.	Nonref.	Ref.	Nonref.
Total students	190	419	190	419
20/100 or poorer	0	0	1	0
20/50 to 20/70	4	0	11	1
20/40	5	0	6	1
20/35	6	1	(1)	(1)
20/30	25	1	34	10
20/25	16	5	(1)	(1)
20/22	21	24	(1)	(1)
20/20	26	82	64	103
20/19	55	151	(1)	(1)
20/17	15	101	(1)	(1)
20/15 or better	17	54	74	304

[1] Score value not on the scale as it was used.

TABLE 17

FAR ACUITY, MONOCULAR

First Grade

Cutoff points for the various standards for referral are shown as follows: Study std.: _____; Manufacturer's std.: .....; High std.: – – –; Low std.: –.–.–.–.

	Clinical Examination		Telebinocular		Massachusetts (Snellen)	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	119	420	119	420	119	420
Right eye: Poorer than 20/20	98	152	71	87	37	18
Right eye: 20/20 or better	21	268	48	333	82	402
Left eye: Poorer than 20/20	96	149	71	91	42	28
Left eye: 20/20 or better	23	271	48	329	77	392
Right and Left Eyes Combined						
Total measurements	238	840	238	840	238	840
20/100 or poorer	18	1	27	5	(1)	(1)
20/50 to 20/70	77	18	16	6	(1)	(1)
20/40	(1)	(1)	21	16	[2] 30	[2] 4
20/35	(1)	(1)	(1)	(1)	(1)	(1)
20/30	99	282	(1)	(1)	49	42
20/28	(1)	(1)	24	28	(1)	(1)
20/25	(1)	(1)	20	35	(1)	(1)
20/22	(1)	(1)	34	88	(1)	(1)
20/20	41	512	30	187	159	794
20/19	(1)	(1)	(1)	(1)	(1)	(1)
20/18	(1)	(1)	38	208	(1)	(1)
20/17	(1)	(1)	26	220	(1)	(1)
20/15 or better	3	27	2	47	(1)	(1)

[1] Score value not on the scale as it was used.

[2] Poorer than 20/30.

tained to show their usual distribution, since the other methods of measurement of near point heterophoria were adequate for diagnostic purposes.

The detailed distributions shown for measurements of monocular visual acuity are those for right eye and left eye combined, so there are 1,218 measurements on 609 sixth-grade students and 1,078 measurements

TABLE 18

NEAR ACUITY, MONOCULAR

Sixth Grade

Cutoff points for the various standards for referral are shown as follows: Study std.: ———; Manufacturer's std.:.....; High std.: – – –; Low std.: –.–.–.–.

	Ortho-Rater		Sight-Screener		Telebinocular		Near Vision Test	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	190	419	190	419	190	419	190	419
Right eye: Poorer than 20/20	69	75	70	75	107	142	55	24
Right eye: 20/20 or better	121	344	120	344	83	277	135	395
Left eye: Poorer than 20/20	76	56	81	82	101	139	55	24
Left eye: 20/20 or better	114	363	109	337	89	280	135	395
Right and Left Eyes Combined								
Total	380	838	380	838	380	838	380	838
20/100 or poorer	10	2	6	2	4	1		
20/50 to 20/70	6	1	19	3	9	3	[2] 27	[2] 8
20/40	4	0	14	4	7	0		
20/33 to 20/35	12	3	[1]	[1]	8	1	23	0
20/28 to 20/30	16	10	112	148	12	2	[1]	[1]
20/25	22	20	[1]	[1]	80	112	60	40
20/22	75	95	[1]	[1]	88	162	[1]	[1]
20/20	56	170	141	337	84	274	270	790
20/19	88	220	[1]	[1]	[1]	[1]		
20/18	[1]	[1]	[1]	[1]	42	165		
20/17	77	242	[1]	[1]	31	85		
20/15 or better	14	75	88	344	15	33		

[1] Score value not on the scale as it was used.

[2] Poorer than 20/33.

on 539 first-grade students. To permit comparison of the right and left eyes, the tables also show for each eye the number of scores of 20/20 or better and the number of scores poorer than 20/20.

For some of the measurements in the Sight-Screener procedure it was necessary to choose between 2 possible methods of scoring.

Instructions for administration of the Sight-Screener tests of lateral heterophoria are that the testee is to report where he sees the arrow when he first looks at the test target, and then to tell where he sees it after the

arrow "stops moving." Both scores are recorded. Only the second score has been used in this study, since it was found to have higher test-retest reliability than the first score. (First score: far point, 0.61; near point, 0.67. Second score: far point, 0.77; near point, 0.82).

Manufacturers of the Sight Screener also suggest 2 possible methods of scoring the tests of stereopsis: (1) the best performance obtained regardless of previous error, or (2) the best performance without previous error. For the study the scores used are those for best performance without previous error, since this gave higher test-retest correlations (0.46 and 0.64, respectively for far and near, as compared with 0.44 and 0.52 for best performance regardless of previous error.)

Tables 19 through 36 conclude the Appendix without further text. The concluding Appendix begins on page 89.

TABLE 19

NEAR ACUITY, BINOCULAR

Sixth Grade

Cutoff points for the various standards for referral are shown as follows: Study std.: ———; Manufacturer's std.: .....; High std.: – – –; Low std.: –.–.–.–.

	Ortho-Rater		Sight-Screener		Telebinocular		Near Vision Test	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	190	419	190	419	190	419	190	419
20/100 or poorer	0	0	1	1	0	0		
20/50 to 20/70	2	0	3	0	1	0	[2] 6	[2] 0
20/40	3	0	9	1	0	1		
20/33 to 20/35	3	0	([1])	([1])	4	0	7	2
20/28 to 20/30	4	4	25	23	5	0	([1])	([1])
					.....		–.–.–.–.–	
20/25	16	4	([1])	([1])	17	21	24	7
20/22	31	30	([1])	([1])	41	50	([1])	([1])
							– – – – –	
20/20	55	127	88	160	55	116	153	410
20/19	41	140	([1])	([1])	([1])	([1])		
20/18	([1])	([1])	([1])	([1])	34	126		
20/17	29	93	([1])	([1])	16	58		
20/15 or better	6	21	64	234	17	47		

[1] Score value not on the scale as it was used.

[2] Poorer than 20/33.

TABLE 20

NEAR ACUITY, MONOCULAR AND BINOCULAR

First Grade

Cutoff points for the various standards for referral are shown as follows: Study std.: ———; Manufacturer's std.: .....; High std.: – – –; Low std.: –.–.–.

	Monocular				*Binocular*			
	Telebinocular		Near Vision Test		Telebinocular		Near Vision Test	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	119	420	119	420				
Right eye: Poorer than 20/20	106	263	49	31				
Right eye: 20/20 or better	13	157	70	389				
Left eye: Poorer than 20/20	101	268	48	32				
Left eye: 20/20 or better	18	152	71	388				
Right and Left Eyes Combined								
Total measurements	238	840	238	840	119	420	119	420
20/100 or poorer	3	1	(1)	(1)	0	0	(1)	(1)
20/50 to 20/70	17	2	(1)	(1)	2	1	(1)	(1)
20/40	13	0	[2] 20	[2] 4	4	0	[2] 7	[2] 2
20/33	14	4	(1)	(1)	4	0	(1)	(1)
20/28 to 20/30	12	18	14	3	3	0	6	0
	.....		–.–.–.–.–		.....		–.–.–.–.–.–	
20/25	90	203	63	56	41	43	17	12
20/22	58	303	(1)	(1)	39	152	(1)	(1)
			– –	– –			– – – – – –	
20/20	22	211	141	777	15	149	89	406
20/18	8	78	(1)	(1)	9	54	(1)	(1)
20/17	1	17	(1)	(1)	2	12	(1)	(1)
20/15 or better	0	3	(1)	(1)	0	9	(1)	(1)

[1] Score value not on scale as it was used.

[2] Poorer than 20/30.

TABLE 21

FAR LATERAL HETEROPHORIA

Sixth Grade

Cutoff points for the standards for referral are shown as follows: Study std.: _____; Manufacturer's std.: .....

Prism diopters of heterophoria	Clinical Examination				Ortho-Rater		Sight-Screener		Tele-binocular		Massachusetts	
	Rod		Cover									
	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.
Total	190	419	190	419	190	419	190	419	190	419	190	419
No reading	9	2	0	1	17	11	4	0	6	0	1	0
Exophoria:												
15 or more	1	0	4	0	(1)	(1)	0	0	1	1		
13–14	0	0	0	0	(1)	(1)	2	1	1	1		
11–12	1	1	0	0	(1)	(1)	0	1	0	0		
9–10	1	1	1	1	(1)	(1)	4	1	(1)	(1)		
8	0	0	1	0	(1)	(1)	0	0	2	2		
7	1	0	0	0	(1)	(1)	2	1	(1)	(1)		
6	2	2	0	0	1	1	0	1	7	14		
5	1	1	2	1	0	0	5	9	(1)	(1)		
4	0	2	0	1	1	2	5	12	(1)	(1)	[2]4	[2]3
3	8	8	1	1	1	0	21	51	39	83		
2	10	27	7	19	3	5	20	58	(1)	(1)		
1	23	70	29	78	7	12	40	117	55	151		
Orthophoria	23	78	4	4	16	47	23	68	(1)	(1)	[2]163	[2]404
Esophoria:												
1	36	113	98	257	26	71	23	50	(1)	(1)		
2	24	65	18	47	29	95	13	24	38	110		
3	15	21	7	6	26	66	9	14	(1)	(1)		
4	8	12	6	2	9	49	3	3	24	46		
5	4	5	0	0	21	27	5	6	(1)	(1)		
6	5	3	1	0	9	13	1	0	(1)	(1)		
7	1	1	1	0	16	9	4	2	8	9	[2]22	[2]12
8	3	0	2	0	8	11	0	0	(1)	(1)		
9–10	5	2	2	0	(1)	(1)	1	0	3	2		
11–12	3	2	1	0	(1)	(1)	3	0	4	0		
13–14	2	2	0	0	(1)	(1)	2	0	2	0		
15 or more	4	1	5	1	(1)	(1)	(1)	(1)	0	0		

[1] Score value not on scale as it was used.

[2] Massachusetts Vision Test score equivalents are:
4 p. d. or more of exophoria;
Less than 4 p. d. of exophoria and 6 p. d. of esophoria;
6 p. d. or more of esophoria.

TABLE 22

NEAR LATERAL HETEROPHORIA

Sixth Grade

Cutoff points for the standards for referral are shown as follows: Study std.: _____; Manufacturer's std.:.....

Priam diopters of Heterophoria	Clinical Examination						Ortho-Rater		Sight-Screener		Tele-binocular		Massa-chusetts	
	Rod		Wing		Cover									
	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.	Ref.	Non-ref.
Total	190	419	190	419	190	419	190	419	190	419	190	419	190	419
No reading	9	2	19*	44*	0	1	7	3	4	0	0	0	1	0
Exophoria:														
15 or more	1	1	1	0	7	0	(1)	(1)	0	0	0	0		
13–14	0	0	0	0	1	0	(1)	(1)	3	0	(1)	(1)		
11–12	2	0	1	1	0	5	1	2	6	4	2	0		
10	1	2	1	0	1	0	(1)	(1)	2	1	(1)	(1)		
9	2	4	0	1	0	0	4	0	6	6	(1)	(1)		
8	3	7	1	1	2	4	2	1	2	4	8	10	[2] 10	[2] 5
7	7	8	1	2	1	1	(1)	(1)	9	19	(1)	(1)		
6	4	6	5	12	4	6	4	3	4	9	(1)	(1)		
5	10	19	3	6	7	11	5	7	18	50	20	18		
4	11	27	22	44	17	30	(1)	(1)	9	28	(1)	(1)		

3	15	42	10	27	23	53	7	11	24	79	(1)	(1)		
2	30	75	34	86	32	74	7	21	11	34	(1)	(1)		
1	25	65	40	97	50	120	(1)	(1)	32	75	43	59		
Orthophoria	19	48	11	28	2	4	14	41	15	34	(1)	(1)	[2] 167	[2] 404
Esophoria:														
1	13	34	14	37	15	64	(1)	(1)	12	38	(1)	(1)		
2	9	22	8	18	5	21	21	58	3	8	(1)	(1)		
3	6	18	3	4	5	10	27	102	10	14	46	190		
4	4	14	5	5	3	5	(1)	(1)	3	3	(1)	(1)		
5	3	7	3	3	2	5	37	69	4	8	(1)	(1)		
6	2	5	2	1	2	2	16	46	1	1	(1)	(1)	[2] 12	[2] 10
7	2	4	2	1	1	0	(1)	(1)	2	3	37	106		
8	1	1	3	1	0	1	15	23	1	0	(1)	(1)		
9–10	6	3	0	0	1	1	11	14	4	1	26	33		
11–12	1	1	0	0	3	0	12	18	4	0	(1)	(1)		
13–14	2	2	0	0	1	0	(1)	(1)	1	0	8	3		
15 or more	2	2	1	0	5	1	(1)	(1)	(1)	(1)	(1)	(1)		

* Includes students not tested; see text.

[1] Score value not on the scale as it was used.

[2] Massachusetts Vision Test score equivalents are:
8 p. d. of exophoria or more;
Less than 8 p. d. exophoria and 6. p. d. esophoria;
6 p. d. of esophoria or more.

TABLE 23

FAR LATERAL HETEROPHORIA

First Grade

Cutoff points for the standards for referral are shown as follows: Study std.: _____; Manufacturer's std.: .....

Prism diopters of heterophoria	Clinical Examination				Telebinocular		Massachusetts	
	Rod		Cover					
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	119	420	119	420	119	420	119	420
No reading	15	5	1	1	1	1		
Exophoria:								
15 or more	2	0	6	0	2	0		
13–14	1	0	0	0	0	0		
11–12	0	0	3	0	0	0		
9–10	1	0	0	0	[1]	[1]		
8	1	0	1	0	4	4		
7	1	3	0	0	[1]	[1]		
6	0	0	1	0	5	16		
5	4	3	1	0	[1]	[1]		
4	0	1	2	2	[1]	[1]	[2] 3	[2] 1
3	2	10	1	5	29	178		
2	4	22	6	29	[1]	[1]		
1	9	57	11	57	22	127		
Orthophoria	18	83	2	12	[1]	[1]	[2] 113	[2] 416
Esophoria:								
1	19	104	44	195	[1]	[1]		
2	16	81	26	112	34	76		
3	6	26	3	5	[1]	[1]		
4	5	9	3	2	7	11		
5	3	6	0	0	[1]	[1]		
6	3	3	1	0	[1]	[1]	[2] 3	[2] 3
7	1	0	0	0	5	5		
8	2	4	1	0	[1]	[1]		
9–10	1	2	0	0	3	0		
11–12	2	0	1	0	2	0		
13–14	0	1	0	0	2	0		
15 or more	3	0	5	0	3	2		

[1] Score value not on scale as it was used.

[2] Massachusetts Vision Test score equivalents are as footnoted in table 21.

TABLE 24

NEAR LATERAL HETEROPHORIA

First Grade

Cutoff points for the standards for referral are shown as follows: Study std.: _____; Manufacturer's std.:.....

Prism diopters of heterophoria	Clinical Examination						Tele-binocular		Massachusetts	
	Rod		Wing		Cover					
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	119	420	119	420	119	420	119	420	119	420
No reading	18	7	86*	295*	1	1	1	0		
Exophoria:										
15 or more	3	0	0	1	4	0	4	0		
13–14	0	0	0	0	0	0	[1]	[1]		
11–12	3	2	0	0	3	1	1	0		
10	1	1	1	0	0	0	[1]	[1]		
9	2	1	0	0	1	0	[1]	[1]		
8	1	3	1	2	2	6	4	7	[2] 3	[2] 2
7	1	5	0	0	0	2	[1]	[1]		
6	1	9	1	1	3	4	[1]	[1]		
							.....			
5	2	13	0	1	2	3	21	69		
4	3	25	3	26	7	19	[1]	[1]		
3	9	38	2	4	10	35	[1]	[1]		
2	18	72	14	46	17	91	[1]	[1]		
1	14	56	6	20	16	98	28	130		
Orthophoria	10	65	1	5	1	8	[1]	[1]	[2] 111	[2] 414
Esophoria:										
1	10	40	1	9	23	83	[1]	[1]		
2	8	31	1	5	12	42	[1]	[1]		
3	5	20	1	1	5	18	41	180		
							.....			
4	2	17	0	2	1	6	[1]	[1]		
5	1	3	0	1	1	1	[1]	[1]		
6	2	6	1	1	2	1	[1]	[1]	[2] 5	[2] 4
7	1	1	0	0	0	0	11	27		
8	1	2	0	0	1	1	[1]	[1]		
9–10	1	1	0	0	1	0	8	6		
11–12	1	2	0	0	0	0	[1]	[1]		
13–14	0	0	0	0	0	0	0	1		
15 or more	1	0	0	0	6	0	[1]	[1]		

* Includes students not tested; see text.

[1] Score value not on the scale as it was used.

[2] Massachusetts Vision Test score equivalents are as footnoted in table 22.

TABLE 25

FAR VERTICAL HETEROPHORIA

Sixth Grade

Cutoff points for the standards for referral are shown as follows: Study std.:_____; Manufacturer's std.: .....

Prism diopters of hyperphoria	Clinical Examination				Ortho-Rater		Sight-Screener		Tele-binocular		Massachusetts	
	Rod		Cover									
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total.....	190	419	190	419	190	419	190	419	190	419	190	419
No reading.....	9	2	0	1	5	2	5	1	3	0	1	0
L. more than 2.0....	0	0	4	0	[1]	[1]	[1]	[1]	[1]	[1]		
L. 1.6–2.0.....	3	0	0	0	1	1	0	0	[1]	[1]		
L. 1.1–1.5.....	0	0	0	0	3	1	0	0	1	0	[2] 7	[2] 1
L. 0.6–1.0.....	10	11	1	0	9	11	3	1	8	1		
L. 0.3–0.5.....	0	0	0	0	45	71	13	22	14	25		
L. 0.2–R. 0.2.....	153	398	180	415	92	293	112	302	156	387	[2] 176	[2] 413
R. 0.3–0.5.....	0	0	0	0	18	16	52	88	5	4		
R. 0.6–1.0.....	9	6	3	2	5	9	2	3	2	2		
R. 1.1–1.5.....	1	0	0	0	12	15	3	2	1	0	[2] 6	[2] 5
R. 1.6–2.0.....	4	0	1	0	[1]	[1]	[1]	[1]	[1]	[1]		
R. more than 2.0...	1	2	1	1	[1]	[1]	[1]	[1]	[1]	[1]		

[1] Score value not on the scale as it was used.

[2] Massachusetts Vision Test score equivalents are:
1.25 p. d. or more of left hyperphoria;
Less than 1.25 p. d. of hyperphoria;
1.25 p. d. or more of right hyperphoria.

TABLE 26

FAR VERTICAL HETEROPHORIA

First grade

Cutoff points for the standards for referral are shown as follows: Study std.: ——; Manufacturer's std.: .....

Prism diopters of hyperphoria	Clinical Examination				Tele-binocular		Massachusetts	
	Rod		Cover					
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	119	420	119	420	119	420	119	420
No reading	17	5	1	1	1	1	0	0
Left more than 2.0	2	1	3	1	(1)	(1)		
Left 1.6–2.0	2	2	0	1	(1)	(1)		
Left 1.1–1.5	0	0	0	0	1	0	[2] 3	[2] 0
Left 0.6–1.0	2	2	1	0	0	1		
Left 0.3–0.5	0	0	0	0	1	0		
Left 0.2—right 0.2	94	401	113	416	112	409	[2] 111	[2] 417
Right 0.3–0.5	0	0	0	0	2	6		
Right 0.6–1.0	0	5	0	0	2	3		
Right 1.1–1.5	0	0	0	0	0	0	[2] 5	[2] 3
Right 1.6–2.0	1	3	0	1	(1)	(1)		
Right more than 2.0	1	1	1	0	(1)	(1)		

[1] Score value not on the scale as it was used.

[2] Massachusetts Vision Test score equivalents are as footnoted in table 25.

TABLE 27

NEAR VERTICAL HETEROPHORIA

Sixth Grade

Cutoff point for Study standard for referral is shown by line across columns.

Prism diopters of hyperphoria	Clinical Examination						Ortho-Rater		Sight-Screener	
	Rod		Wing		Cover					
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total	190	419	190	419	190	419	190	419	190	419
No reading	9	2	19*	44*	0	1	4	0	5	1
L. more than 2.0	2	0	2	0	2	0	(1)	(1)	(1)	(1)
L. 1.6–2.0	6	2	5	1	2	0	3	6	2	0
L. 1.1–1.5	0	0	0	0	0	0	5	6	1	0
L. 0.6–1.0	8	12	2	10	2	0	24	38	0	1
L. 0.3–0.5	0	0	0	0	0	0	64	175	16	16
L. 0.2–R. 0.2	156	391	156	361	179	415	66	174	117	314
R. 0.3–0.5	0	0	0	0	0	0	5	3	46	83
R. 0.6–1.0	4	5	4	2	3	1	1	6	2	3
R. 1.1–1.5	0	0	0	0	0	0	18	11	1	1
R. 1.6–2.0	4	5	2	1	0	1	(1)	(1)	(1)	(1)
R. more than 2.0	1	2	0	0	2	1	(1)	(1)	(1)	(1)

* Includes students not tested; see text.

[1] Score value not on the scale as it was used.

TABLE 28

NEAR VERTICAL HETEROPHORIA

First Grade

Cutoff point for Study standard for referral is shown by line across column.

Prism diopters of hyperphoria	Clinical Examination					
	Rod		Wing		Cover	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total	119	420	119	420	119	420
No reading	18	6	86*	295*	1	1
Eye and prism diopters:						
Left more than 2	0	0	1	0	3	2
Left 2	4	5	0	0	1	0
Left 1	1	3	0	0	0	1
0	94	397	32	123	111	415
Right 1	0	5	0	1	0	1
Right 2	1	4	0	1	0	0
Right more than 2	1	0	0	0	3	0

* Includes students not tested; see text.

TABLE 29

DEPTH PERCEPTION

Sixth Grade

Cutoff points for standards for referral are shown as follows: Study std.:_____ Manufacturer's std.:.....

Ortho-Rater (Far point)			Telebinocular (Far point)			Sight-Screener (Far point)			Sight-Screener (Near point)		
Scores	Ref.	Nonref.	Scores	Ref.	Nonref.	Scores	Ref.	Nonref.	Scores	Ref.	Nonref.
	190	419		190	419		190	419		190	419
None	16	5	None	7	1	None	31	11	None	24	6
Trial	29	5	1	5	3						
			2	1	2						
			3	5	0	A	12	3	A	13	6
			4	0	0						
A	13	7	5	5	1						
						B	3	0	B	2	3
B	14	8	6	2	0						
C	20	23	7	1	0						
						C	28	11	C	13	4
D	22	43	8	1	0						
			.....								
E	26	90	9	2	1						
						D	15	28	D	23	20
F	13	65	10	16	6						
G	6	18	11	9	7						
H	31	155	12	136	398	E	101	366	E	115	380

TABLE 30

DEPTH PERCEPTION

First Grade

Cutoff points for standards for referral are shown as follows: Study std.:_____; Manufacturer's std.: .....

Telebinocular (Far point)		
Scores	Ref.	Nonref.
Total	119	420
None	18	10
1	4	8
2	6	9
3	5	10
4	2	9
5	6	5
6	0	6
7	3	2
8	1	2
.....		
9	2	4
10	12	43
11	12	29
12	48	283

TABLE 31

FUSION (TELEBINOCULAR)

Both Grades

Cutoff point for standard for referral is shown by line across column.

	Sixth Grade				First Grade			
	Far Point		Near Point		Far Point		Near Point	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total	190	419	190	419	119	420	119	420
Sees 2	3	0	2	0	4	1	3	0
Underconvergence:								
Sees 4	4	2	2	3	5	2	3	3
Sees 4, then 3	2	3	1	1	3	5	3	4
Sees 3	118	296	129	294	76	350	88	336
Overconvergence:								
Sees 4, then 3	41	99	29	100	28	52	14	62
Sees 4	22	19	27	21	3	10	8	15

TABLE 32

BINOCULAR VISION

Both Grades

Cutoff point for standard for referral is shown by line across column.

	Sight-Screener Sixth Grade					Telebinocular	
	Far Point		Near Point			Far Point	
	Ref.	Nonref.	Ref.	Nonref.		Sixth Grade	First Grade
Total	190	419	190	419	Total	609	539
Reads none	3	0	1	0	Sees nothing	0	0
Right eye only	10	2	7	2	Right eye only	1	0
Left eye only	9	3	7	2	Left eye only	2	1
Alternates	1	1	1	1	Underconvergence	20	4
					Overconvergence	17	4
Binocular	167	413	174	414	Binocular	569	530

TABLE 33

PRISM CONVERGENCE

Both Grades

Break point in prism diopters	Clinical Examination							
	Far				Near			
	6th Grade		1st Grade		6th Grade		1st Grade	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	172	385	53	219	172	385	53	218
3–4	0	0	1	0	2	0	0	0
5–6	0	0	0	0	0	0	1	0
7–8	4	3	0	5	0	0	0	0
9–10	0	7	2	9	0	0	0	1
11–12	14	25	8	27	0	0	0	1
13–14	7	27	5	30	0	0	0	0
15–16	15	38	7	36	1	3	0	0
17–18	20	44	3	23	2	2	0	0
19–20	19	40	8	23	1	4	0	5
21–22	16	48	6	18	7	8	0	0
23–24	10	29	2	9	5	18	0	10
25–26	7	27	2	9	13	14	3	10
27–28	8	18	1	3	10	33	7	20
29–30	52	79	8	27	131	303	42	171

TABLE 34

PRISM DIVERGENCE

Both Grades

Break point in prism diopters	Clinical Examination							
	Far				Near			
	6th Grade		1st Grade		6th Grade		1st Grade	
	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.	Ref.	Nonref.
Total students	171	385	53	218	171	385	53	218
29–30 or more	2	0	0	1	4	6	0	1
27–28	0	0	0	0	7	5	0	2
25–26	0	0	0	0	14	21	3	5
23–24	0	0	0	0	19	28	3	28
21–22	1	1	0	0	26	58	11	51
19–20	1	3	0	0	40	74	14	51
17–18	2	2	0	1	29	102	12	43
15–16	1	1	1	8	11	48	7	26
13–14	3	12	2	5	13	31	3	9
11–12	15	21	5	17	4	9	0	2
9–10	29	48	14	43	2	3	0	0
7–8	50	125	16	70	1	0	0	0
5–6	47	125	12	62	0	0	0	0
3–4	19	46	3	11	0	0	0	0
1–2	1	1	0	0	1	0	0	0

TABLE 35

NEAR POINT CONVERGENCE

Both Grades

Millimeters	Clinical Examination			
	6th Grade		1st Grade	
	Ref.	Nonref.	Ref.	Nonref.
Total students	190	419	119	420
No reading	0	0	2	1
1–10	29	69	32	134
11–20	46	116	47	167
21–30	46	104	21	74
31–40	26	58	7	28
41–50	9	26	1	10
51–75	10	25	6	4
76–100	9	17	2	2
More than 100	15	4	1	0

TABLE 36

SPHERICAL EQUIVALENT

Both Grades

Cutoff points for standard for referral are shown by line across column and footnote.

Clinical Examination					
	Sixth Grade			First Grade	
	Ref.	Nonref.		Ref.	Nonref.
Total students	190	419		119	420
Right eye:			Right eye:		
No reading	1	0			
−0.25 d. or less	51	5	−0.25 d. or less	16	1
0.00 to +3.00 d	118	413	0.00 to +3.50 d	75	419
+3.25 d. or more	20	1	+3.75 d. or more	28	0
Left eye:			Left eye:		
−0.25 d. or less	54	6	−0.25 d. or less	14	0
0.00 to +3.00 d	113	412	0.00 to +3.50 d	82	420
+3.25 d. or more	23	1	+3.75 d. or more	23	0
Total measurements	380	838		238	840
No reading	1	0		0	0
Diopters:					
−4.00 or less	8	0		3	0
−3.00 to −3.75	1	0		0	0
−2.00 to −2.75	13	0		5	0
−1.00 to −1.75	27	0		7	0
−0.25 to −0.75	56	11		15	1
0.00	15	13		2	10
+0.25 to +0.75	83	481		34	227
+1.00 to +1.75	82	312		68	506
+2.00 to +2.75	40	18		32	90
+3.00 to +3.75	[1] 32	[1] 1		[1] 35	[1] 6
+4.00 to +4.75	8	0		20	0
+5.00 to +5.75	6	0		14	0
+6.00 or more	8	2		3	0

[1] Cutoff point is +3.00 for sixth grade, +3.50 for first grade.

APPENDIX D

Discussion of Correlations

The computation of the point correlation coefficient can be illustrated with the values for sixth grade Teacher's Judgment:

		R	N
	609	190	419
R....	197	80	117
N....	412	110	302

We first multiply the numbers in the 2 categories where all students would have fallen if the screening procedure were ideal: $80 \times 302 = 24{,}160$. We next multiply the numbers in the 2 "odd" categories (students missed and over-referrals): $110 \times 117 = 12{,}870$.

Clearly, if there is good agreement between screening procedure and criterion, the first of these products will be much larger than the second. If there were perfect agreement there would be no students in the odd categories and the second product would be zero. The degree of correspondence is indicated by the ***size of the difference between the cross products***, which, in this case, is 11,290.

A further step is required to reduce this figure to a value that is independent of the number of students tested and the proportions of total referrals or nonreferrals by ophthalmologist or screening procedure. This is accomplished by dividing into 11,290 the square root of the product of the marginal totals: $\sqrt{412 \times 197 \times 190 \times 419}$, or 80,383. The quotient $\frac{11{,}290}{80{,}383}$, or 0.14, is the correlation coefficient. This value shows that, insofar as there is any agreement at all between teacher's judgment and the criterion, the agreement is in the "positive" or expected direction; it is worth noting incidentally that without use of the correlation coefficient one could as easily suppose the scatter showed some negative or inverse correspondence between procedure and criterion.

By setting up several scatters and noting the values taken by numerator and denominator, the relation between them is easily seen. For example, if no students fall in the "odd" categories, the denominator will be the same as the numerator, so the coefficient will be 1.00, indicating perfect

agreement. If any students fall in the "odd" categories, the denominator will be greater than the numerator, so the coefficient will be less than 1.00, indicating less than perfect agreement. If there is no difference between the cross products, the numerator will be zero and the coefficient zero, indicating no correlation. If the first cross-product is less than the second, the difference will be a minus value, giving a negative correlation coefficient indicative of inverse correspondence.

The correlation for a 2 x 2 scatter may be computed either as shown above or by the method ordinarily used for larger scatters. The procedure shown above is easier with, and is applicable only to, a 2 x 2 scatter, but the result is a true "product-moment" coefficient and is identical with the result obtained by the other procedure.

The correlation computed directly from a 2 x 2 scatter is called a point coefficient because its use involves the assumption that the referral and nonreferral categories are simply "either-or" or "point" distributions. In computing the correlations on the basis of this assumption we understate the level of agreement that would be found if we could use a full distribution of screening scores and a full distribution of ophthalmologist's scores in computing each correlation.

However, the understatement of the agreement is similar for the various screening procedures, and the amount of understatement is not great where, as in this study, the general level of correspondence between the procedures and the criterion is not high.

So far as it is possible to take account of underlying distributions, and thus to "correct" for coarse grouping into the referral and nonreferral categories, the tetrachoric correlation coefficients in table 37 provide the best available solution. These coefficients involve the assumption that the distributions of scores underlying the referral and nonreferral categories conform to the "normal" distribution. Since this is by no means certain, the relatively high values of these coefficients may be misleading as regards the general level of agreement between screening scores and criterion scores. Otherwise, use of the tetrachoric values in place of the point coefficients in table 2b would have little effect on conclusions of the study.

We have not used "contingency coefficients" because they would understate the agreement between screening scores and criterion scores even more than the point correlations. If desired, however, the contingency coefficient may be obtained for any scatter by taking the square root of $r^2/(1+r^2)$,where r is the value of the point correlation. Also, chi-square may be found by multiplying the number of cases (609 or 606 for the grade concerned) by the square of the point correlation. The tetrachoric coefficient has no simple relationship to the point correlation, the contingency coefficient, or chi-square.

The statement is made in the text (p. 22) that there is no rigorous way to test the statistical significance of the differences among the correlation coefficients. In theory, at least, such tests would be possible through the

TABLE 37

Tetrachoric Correlations

Each correlation is for referral or nonreferral by the given screening procedure vs. referral or nonreferral by the ophthalmologist. For the point coefficients analogous to these tetrachoric coefficients, see the column headed "Correlations" in table 2b.

Test	Tester	Correlation	
		Sixth grade	First grade
Judgment	Teacher	0.24	0.30
Snellen, high std	Technician	.82	.67
	Nurse	67	.68
	Teacher	.70	.70
Snellen, low std	Technician	.73	.68
	Nurse	.69	.80
	Teacher	.76	.68
Massachusetts	Technician	.72	.63
	Nurse	.63	.63
Ortho-Rater	Technician	.53	(1)
	Nurse	.53	(1)
Sight-Screener	Technician	.63	(1)
	Nurse	.44	(1)
Telebinocular, study std	Technician	.52	.61
	Nurse	.46	.38
Telebinocular, mfr's. std	Technician	.46	.51
	Nurse	.30	.38
Judgment and Snellen, high std	Teacher	.48	.52
Judgment and Massachusetts	Teacher and Technician [2]	.50	.50
	Teacher and Nurse [2]	.50	.50
Near Vision, high std	Technician	.51	.68
	Nurse	.46	.45
Near Vision, low std	Technician	.64	.65
	Nurse	.63	.55
Snellen, high std., and Near Vision, high std.	Technician	.72	.64
	Teacher and Nurse [3]	.63	.53
Snellen, high std., and Near Vision, low std.	Technician	.82	.67
	Teacher and Nurse [3]	.72	.73
Judgment, Snellen high std., and Near Vision, high std.	Teacher and Nurse [4]	.50	.48

[1] Not administered to first grade.
[2] Judgment by Teacher.
[3] Snellen by Teacher, Near Vision Test by Nurse.
[4] Judgment and Snellen by Teacher, Near Vision Test by Nurse.

use of further scatters for the numerous intercorrelations of the screening procedures. But the tests could not be made and interpreted without involving some arbitrary assumptions about the distributions underlying the referral and nonreferral categories.

The study shows that all of the correlations of screening procedures with the criterion are low, and there is reason to think that they can and will be improved. Since little relationship is to be expected between the efficiency of the procedures as found in this study and the efficiency they may show after improvement, the worthwhileness of testing the statistical significance of differences among the procedures as they stand may be doubted.

BOSTON PUBLIC LIBRARY

3 9999 05708 6587

U. S. GOVERNMENT PRINTING OFFICE: 1954

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price 35 cents

CHILDREN'S BUREAU PUBLICATION NUMBER 345 • 1954

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Social Security Administration
Children's Bureau